Denoising of Carpal Bones for Computerised Assessment of Bone Age

Darin S O'Keeffe

Department of Electrical and Computer Engineering

A thesis presented for the degree of
Doctor of Philosophy

University of Canterbury
Christchurch, New Zealand
January 2010

”Do you want to know what knowledge is? When you know something recognise that you know it, and when you don’t know something, recognise that you don’t know it. That’s knowledge.”

The Analects of Confucius, Book 2 Chapter 17.

Contents

Acknowledgements

No one does PhD research on their own. During my studies I had support ranging from as simple as a new video cable for my computer, to pieces of software implementing published methods. A very big thank you to everyone who has helped me. But there are some special acknowledgements that I would like to make.

Firstly, thank you to Professor Philip Bones, my Senior Academic Supervisor. Phil showed great skill, patience, support, and confidence that I could complete my part-time PhD studies. Not only did he help with the technicalities of my research, he always listened and gave an honest opinion. There were plenty of tangents for me to go off on and Phil tried his hardest to make sure I didn't do this.

Dr Michael Hurrell, Associate Supervisor, had the original idea many years ago for an automatic bone age assessment system. At that time a commercial system looked just around the corner, but such a system never eventuated. The desire is still there for a full assessment system, but the part I bit off for my PhD studies was analysis of the carpal bones. Mike always gave constructive criticism and was wonderful with corrections to my writing style. Mike always presented the challenge, "if I can do it, a computer should be able to". There is still a long way to go, but I hope this thesis helps address part of this challenge and paths the way for further local research in this area. Thank you, Mike.

To Dr Richard Tremewan for his long-term support both as an Associate Supervisor and as my manager. I acknowledge that at times this presented a challenge for him and I hope I didn't take too many years off his life.

My employer, the Canterbury District Health Board, helped me get started on a pilot project looking at automatic bone age assessment. Initially this was a small grant for software and books, but this grew to some time off work to allow me to work on my PhD part-time, and some financial assistance with my PhD fees for part of my study period.

I am indebted to the University of Canterbury for granting me a Doctoral Scholarship for part of my PhD studies. This came at a time that was financially challenging and was much

appreciated.

Thank you to my work colleagues who endured my just-in-time management during periods of my studies.Your support has been appreciated.

My dearest thanks go to my boys Alex, Cameron, and Hamish, and *especially* my wife Robyn to whom I owe everything non-academic in this thesis. It is hard to thank you all in words. You gave me time away from my responsibilities as a father and husband, and on top of that you put up with what at times were not my finest moments (sorry). I owe you and I love you.

Irrespective of the people that have helped support my PhD studies, I alone take responsibility for any errors, omissions, or oversights that may be present in this thesis.

Darin O'Keeffe

Abstract

Bone age assessment is a method of assigning a level of biological maturity to a child. It is usually performed either by comparing an x-ray of a child's left hand and wrist with an atlas of known bones, or by analysing specific features of bones such as ratios of width to height, or the degree of overlap with other bones. Both methods of assessment are labour intensive and prone to both inter- and intra-observer variability. This is motivation for developing a computerised method of bone age assessment.

The majority of research and development on computerised bone age assessment has focussed on analysing the bones of the hand. The wrist bones, especially the carpal bones, have received far less attention and have only been analysed in young children in which there is clear separation of the bones. An argument is presented that the evidence for excluding the carpal bones from computerised bone age assessment is weak and that research is required to identify the role of carpal bones in the computerised assessment of bone age for children over eight years of age.

Computerised analysis of the carpal bones in older children is a difficult computer vision problem plagued by radiographic noise, poor image contrast, and especially poor definition of bone contours. Traditional image processing methods such as region growing fail and even the very successful Canny linear edge detector can only find the simplest of bone edges in these images. The field of partial differential equation-based image processing provides some possible solutions to this problem, such as the use of active contour models to impose constraints upon the contour continuity. However, many of these methods require regularisation to achieve unique and stable solutions. An important part of this regularisation is image denoising.

Image denoising was approached through development of a noise model for the Kodak computed radiography system, estimation of noise parameters using a robust estimator of noise per pixel intensity bin, and incorporation of the noise model into a denoising method based on oriented Laplacians. The results for this approach only showed a marginal improvement when using the signal-dependent noise model, although this likely reflects how

the noise characteristics were incorporated into the anisotropic diffusion method, rather than the principle of this approach. Even without the signal-dependent noise term the oriented Laplacians denoising of the hand-wrist radiographs was very effective at removing noise and preserving edges.

Preface

Thesis Outline

Chapter 1
Introduces the different methods of bone age assessment and the need for automation as a means of reducing inter- and intra-observer variability. Background terminology is also introduced in this chapter.

Chapter 2
An extensive review of computer-assisted, user-assisted, and automated methods of estimating bone age. Also presented in this chapter is a model of a system for computerised assessment of bone age. The chapter concludes with a review of the role of the carpal bones and the importance of researching the role of these bones in systems for the computerised assessment bone age. The area that has received the least attention in the computerised assessment is the role of the carpal bones in older children.

Chapter 3
Essential mathematical concepts, terminology, and notation.

Chapter 4
The pixel intensity and noise characteristics of the Kodak CR system is analysed in-depth and a noise model is developed for the radiographic images. The pixel intensity and noise characteristics were studies using a phantom specially constructed to provide a radiograph that allowed the computed radiography system to be testing in the 'clinical' hand processing mode.

Chapter 5
A comprehensive review of preprocessing tasks for computerised bone age assessment. The findings from Chapter 4 are used to identify radiographic markers.

Chapter 6
Impulse noise removal is addressed using a dual-task of detection and elimination. A me-

dian absolute deviation (MAD) robust statistic is used to detect the impulse noise. The same estimator is used to estimate the noise components of the radiographic images and this noise model is used in a denoising method based on anisotropic diffusion. This chapter includes an extensive investigation into the use of the MAD estimator for measuring the statistics of radiographic images that undergo logarithm conversion of the pixel intensities. The chapter concludes with the finding that at this stage incorporation of the noise information into the oriented Laplacians diffusion only marginally improves the success of the denoising method.

Algorithm implementation

Unless stated otherwise, all algorithms and numerical methods used in this thesis were implemented using MATLAB® R2007b, with some function support from the Image Processing Toolbox — The MathWorks Inc. (http://www.mathworks.com, accessed September 2009). Additional functionality was added to MATLAB® using Microsoft Visual Studio© 2005. Sigmaplot® 8.02a (http://www.sigmaplot.com) was used to fit curves to most of the data presented in this thesis. This fitting used the Marquardt-Levenberg algorithm.

Abbreviations and acronyms

Abbreviations and acronyms used in this thesis are listed here for reference . Mathematical notation is introduced in Chapter 3.

2D, 3D	two-dimensional and three dimensional, respectively
ACM	active contour model
CASAS	computer-aided skeletal age scoring system
CR	computed radiography
CROI	carpal region of interest
DICOM	Digital Imaging and Communications in Medicine
EI	Exposure Index – specific to the Kodak CR system
Eqn.	equation
Fels	FELS (bone age assessment method)
GFR	Gibbs random field
GHD	growth hormone deficiency
GP	Greulich–Pyle (bone age assessment method)
Gy	gray – a measure of radiation dose
kVp	peak kilovoltage applied to an x-ray tube
MAD	median absolute deviation - a robust estimator of statistical dispersion
PACS	picture archiving and communication system
PROI	phalangeal region of interest
ROI	region of interest
RMS	root mean square
RUROI	radius-ulna regions of interest
RUS	radius, ulna, and short bones
SNR	signal-to-noise ratio
TW	Tanner–Whitehouse (bone age assessment method)

Chapter 1

Assessment of Skeletal Maturity

The evaluation of skeletal maturity through bone age assessment is a frequently performed paediatric x-ray examination, and is important for the management of childhood growth disorders [Rucc95]. The assessment is based on analysis of features in an x-ray of a child's left hand and wrist, and compares these with known standards to estimate a child's biological age.

Accurate and reproducible bone age assessment is a difficult and time consuming radiological procedure. It is mainly used as part of the assessment of children with growth disorders. This chapter provides background material about the procedure and describes the problems associated with it. Building on the need for improved accuracy and reproducibility of bone age assessment in clinical applications, the chapter finishes with a discussion of the need for automated assessment of bone age.

1.1 Skeletal maturation

Many physiological and anatomical changes occur as a child matures. The most obvious is a change in height that is often used to monitor the growth of children. The skeletal differences between young and old children extend beyond height alone. For example, there are large differences in the shape of some bones as a child ages. It is these bony differences that define skeletal maturity as an indicator of the physical maturity of a child.

This section provides bone development information that is necessary to understand the process of skeletal maturation of the hand and wrist and the concept of bone age. It introduces terminology that is used throughout this thesis.

1.1.1 Skeletal development

The change in height of a child is mostly due to elongation of the long bones of the skeleton. The bones lengthen because of the proliferation of cells in growth plates at the ends of the bone. As these cells develop, bone matrix is laid down and this calcifies to form bone tissue, a process referred to as ossification.

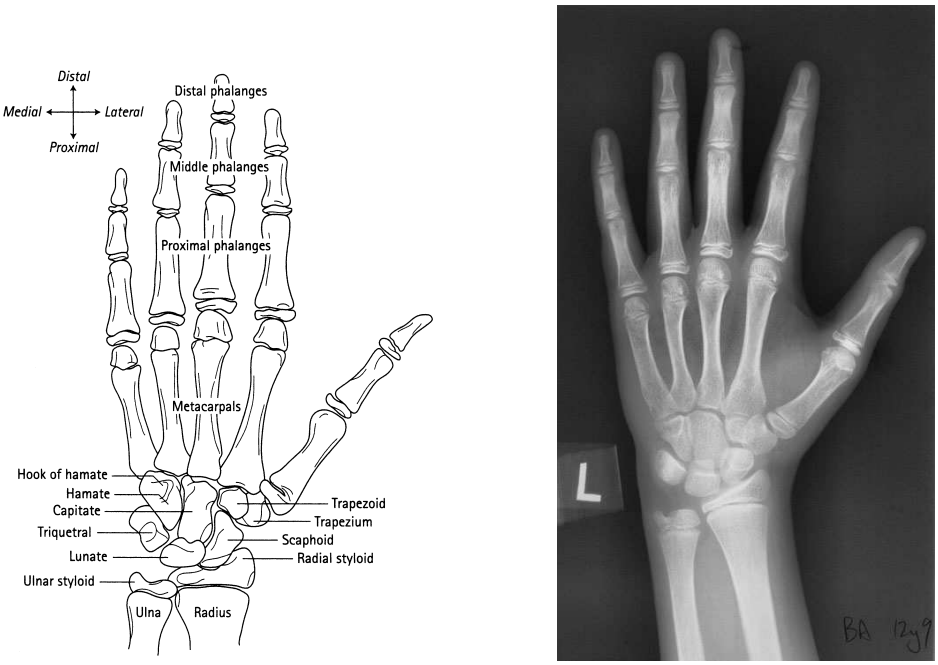
The bones do not just grow longer and wider as a child ages. For example, although many of the bones of the hand are present at birth, the bones of the wrist (the carpal bones) appear at various times during early childhood. Similarly, the ends of some bones appear and develop separately, then fuse with the main part of the bone when growth is complete. It is important to understand these types of changes because many can be demonstrated radiographically and used as indicators of skeletal maturity.

1.1.1.1 Bone development and terminology

There are two types of hand and wrist bones: long bones and short bones. Long bones, as the name suggests, have a dominant long axis. They include the radius, ulna, and phalanges (Figure 1.1(a)). Short bones have no dominant long axis. The short bones of the wrist, the carpal bones, are also shown in Figure 1.1(a) in a group between the metacarpals and the radius and ulna. The two types of bones differ in their development.

The long bones start in the fetus as a model of cartilage surrounded by fibrous tissue. Around the seventh week of fetal development the cells in the middle of the cartilage begin to calcify in an area called the primary ossification centre. At the time of birth most long bones have formed a tubular shaft called a diaphysis. The diaphysis is composed of an outer tube of compact bone and central spongy bone containing a marrow cavity [Spen92, p160]. At each end of the diaphysis is a cartilaginous cap, a region of growth that is growing away from the expanding primary ossification centre. This cartilage later forms its own ossification centre, but at this stage the unossified cap of cartilage or epiphysis is separated from the end of the diaphysis by a section of cartilage called the growth plate or physeal plate (Figure 1.2). The end of the diaphysis nearest the growth plate is called the metaphysis. The region of the growth plate next to the metaphysis is sometimes referred to as the epiphyseo-diaphyseal junction. It is this junction that is responsible for the elongation of the diaphysis and, hence, elongation of the bone.

The short bones of the wrist develop differently from the long bones. Although there are similarities in the stages of ossification, the carpal bones are not ossified at birth and they ossify solely via their primary ossification centres [Roch88, p7]. The carpal bones develop from a cartilage model whose shape resembles what the bone will grow to look like [Roch88, p8]. The cells in a piece of the cartilage become enlarged (hypertrophied) then



(a) (b)

Figure 1.1 Radiographic anatomy of the left hand and wrist (a) outline of bones of the hand and wrist (reproduced with permission from [Tann01] – Copyright Elsevier), (b) a typical radiograph for bone age assessment. The carpal bones are the group between the metacarpals and the radius and ulna

calcify to form spongy bone (cancellous bone) surrounded by a shell of compact bone. The bone enlarges through a complex process resulting in conversion of the cartilage to spongy bone, with the compact bone remaining approximately the same thickness. When the bone approaches full maturity, an outer layer of cartilage remains as an articulating surface for the bone.

Significant variations in carpal development can occur due to varying order of appearance of bones, unexpected fusion or partitioning of bones, and formation of accessory elements from nodules of cartilage. The potential for these variations needs to be considered when investigating methods for incorporating carpal bone development into the assessment of skeletal maturity. Even though these variations exist, there are some predictable components to the carpal bone development. For example, the capitate and hamate bone ossification centres are either the first or second to appear following ossification [Garn60]. These predictable components can assist with the identification of the carpal bones on hand-wrist radiographs.

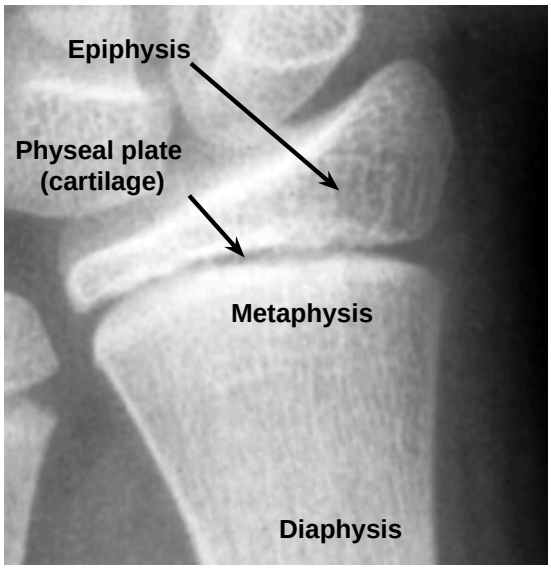


Figure 1.2 Terminology used to describe epiphyseal regions.

1.1.1.2 The epiphysis, growth plate and bone fusion

The bony epiphysis forms as a secondary ossification centre in the cartilaginous cap at the end of the bone. The process of bone formation is essentially the same as the primary ossification centre, but occurs considerably later. The bony epiphysis is initially spherical, so its projection onto a radiograph is circular. As it grows, it becomes more hemispherical with its flat side adjacent to the growth plate [Buck00, p81]. Within a few years, the only cartilage that remains is a thin layer on the outer, articular surface of the bony epiphysis and the growth plate.

The growth plate is composed of cartilage that has three distinct zones, each having a unique structure, biochemistry, and function [Buck00, p81]. The zone nearest the bony epiphysis is called the reserve zone and although it does have the capacity to produce a cartilage cell matrix, its function is largely unknown [Buck00, p82]. The middle, proliferation zone is composed of dividing and matrix-forming cartilage cells responsible for the thickening of the plate. The zone closest to metaphysis/diaphysis is the zone of provisional calcification and is composed of large cartilage cells that die and become calcified. The calcified cells are invaded by cells from the metaphysis that break them down to form new bone tissue. The rate of cell formation in the reserve and proliferation zones, and the

rate of cell loss at the junction with the metaphysis, are usually the same, so the growth plate thickness is maintained as the bone lengthens.

The growth plate widens transversely as the bone grows longitudinally, but the face of the epiphysis that is at the junction with the plate widens more rapidly than either the plate or the metaphysis [Roch88, p6]. The result is a change in the ratio of the metaphysis-epiphysis diameters with maturation. Similarly, a thin radiopaque layer of bone appears on the metaphyseal side of the epiphysis as the metaphysis and epiphysis become nearly parallel prior to growth plate fusion [Roch88, p6]. These are both examples of bony changes that act as indicators of maturation due to changes in the growth plate.

During the late stages of bone maturation, the cartilage growth rate on the epiphysis side of the growth plate slows, causing diminishing growth plate thickness [Buck00, p161]. Starting from the centre of the bone, the epiphysis and diaphysis begin to fuse [Roch88, p7]. This process is sometimes referred to as epiphyseo-diaphyseal fusion, or simply fusion of the growth plate. Once the growth plate has been replaced by bone, the bone cannot grow any longer. All that remains is a thin, radiopaque layer of bone referred to on radiographs as an epiphyseal line. This layer is resorbed after fusion, but the process may take several years [Roch88, p7]. It is at this point that the bone is considered fully mature.

Many factors influence the proliferation and transformation of cells within the growth plate, and the interaction of the growth plate with the metaphysis. For example, growth hormone has an effect on cellular proliferation [Buck00, p90]. Genetic factors, hormones and vitamins affect the growth plate, and imbalances in these growth factors will influence the growth and maturation of the bone. For example, a deficiency of thyroid hormone results in a decrease in cell division in the proliferation zone, causing growth retardation [Buck00, p89]. These bone changes can be investigated radiographically and used as markers of disease.

1.1.2 Growth and skeletal maturity

Assessment of the degree of physical maturity of a child is important for both their medical and psychological management. Surgeons require assessment of the physical maturity of children to plan some orthopaedic operations. Certain medications and therapeutic procedures have an effect upon the growth and development of children, and assessment of physical maturity can influence decision making and patient management. Clinicians frequently face children (and parents) who want to know why they are much taller or shorter than their friends, and if they are going to continue to be that way. Often patient history and clinical observations will be sufficient to offer reassurance, but in many circumstances it is important to estimate physical maturity using a reliable and easy-to-perform proce-

dure.

This section introduces background information on growth and discusses the role of skeletal maturity in the assessment of physical maturity.

1.1.2.1 Height as a maturity indicator

Normal childhood growth is a complex process involving the interaction of numerous intrinsic and extrinsic factors. Although children normally grow in a predictable and orderly manner, some children do not have a normal pattern of growth and this can be the first sign of a number of disease processes [Wils98, p1428]. It is for this reason that accurate measurement of height is an important part of the physical examination of a child [Wils98, p1428] [Gree01, p163]. The first absolute measurement of a child usually gives a deviation from normal childhood development by comparing selected body measurements against a series of percentile curves for a given population of normal children.

These percentile curves have their limitations. Firstly, they are usually developed from cross-sectional data - measurements made at a single point in time, or in this case, age. The curves are an average across the population and are of questionable value for the assessment of adolescents because of age differences for the onset of puberty [Wils98, p1430]. Secondly, the curves do not properly define the growth status of children above the 97th and below the 3rd percentiles; these being the children who require the most thorough assessment. Lastly, a child who has a normal initial height measurement, according to the percentile curves, may not be growing normally. The growth of a tall child may slow down or stop, but they may not fall below normal range of the percentile curves. It is for this reason that subsequent measurements are usually required to determine the child's growth rate or velocity. Finally, some syndromes caused by a genetic abnormalities result in growth rates and patterns that cannot be monitored using growth charts [Roch88, p12]. Such children require assessment of maturity using other means.

1.1.2.2 Maturity and skeletal maturity indicators

The physical development of normal children from soon after birth through to adolescent puberty is usually a predictable and orderly sequence of stages. Although the concept of physical maturity is intuitively simple, defining it and using a method to measure it is difficult [Tann01, p1] [Mars77, p87].The problem is that children pass through these stages at both varying rates and varying ages. This means that children of the same chronological age may differ in both the rate they approach (physical) adulthood, and how far they are away from that final point of development – physical maturity.

Useful indicators of physical maturity should ideally be based on maturational processes

that occur in all normal individuals during growth, and they should have an endpoint that is the same in all individuals [Mars77, p87].

A maturity indicator describing a single event is considered of little value. It is more useful to have a sequence of events or stages that all individuals pass through in the same order [Tann01, p1] [Mars77, p87]. Ideally, this sequence of stages should evenly span the physical development of a child from birth to adulthood, with each stage representing an equal step towards adult maturity [Mars77, p87]. However, no such sequence currently exists. A practical option is to use multiple different overlapping maturational sequences. Because there are many bones of the skeleton that grow and mature at various times during the development of a child, the skeleton provides many simultaneous sequences, each of which can be described by skeletal maturity indicators.

The changes seen in the skeleton are not just growth of the long bones, but include appearance of new growth centres, fusion of existing growth plates, and changes in the shape of bones. These changes can be viewed in their entirety as the successive transformation of bones from initial ossification through to their final shape and growth plate fusion, where relevant. This process is called skeletal maturation. It is a remarkably orderly process and usually involves sequences of events that are reproducible from one child to the next [Zeri91].

Quantifying skeletal maturity is difficult because skeletal maturation is a dynamic, continuous process with no simple measure such as length. Most methods of assessing skeletal maturity use skeletal maturity indicators to artificially divide the developmental interval from newborn to adult. These skeletal maturity indicators are usually based on the plain film radiographic appearance of bone tissue and include ossification centre relative size and shape, and physeal narrowing and ossification.

The types of skeletal maturity indicators depend on the bone being investigated and, hence, on the area of the body being studied. The two most common indicators for bones are the onset of ossification of the bone or epiphyses, and the fusion of the physes (Figure 1.3. These are the two extreme stages of development and there is much more information available in the intermediate stages of bone development between these two points. Any method of skeletal maturity assessment would be inefficient if it was based only on these two indicators, but the difficulty with using intermediate stages is that some of them are hard to recognise and the resulting errors can have a large impact upon any assessment [Mars77, p88]. To be useful, the indicators for intermediate stages should have occurred in all children who have attained a certain level of skeletal maturity. Furthermore, any intermediate stages should allow for a reasonably uniform division of the span of development.

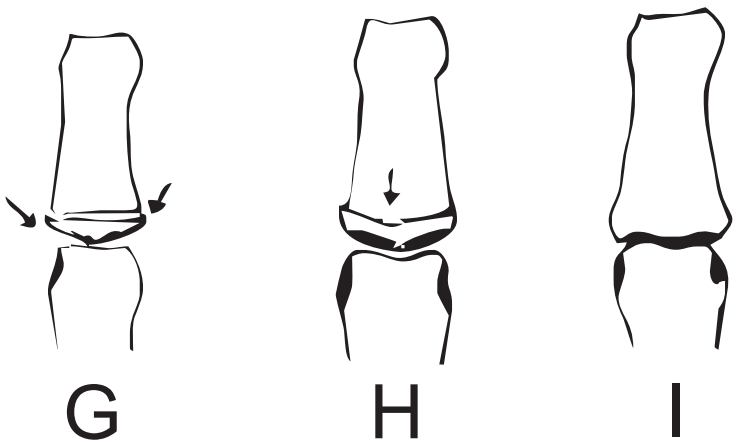


Figure 1.3 Examples of skeletal maturity indicators for the middle phalanges. As the bones mature, the epiphysis caps the metaphysis (G), followed by the start of fusion of the epiphysis and metaphysis (H), and finally the completed fusion of the epiphysis and metaphysis - with the line of fusion disappearing (I). The stages G, H, and I are from the TW3 method introduced in Section 1.2.3.2 (adapted with permission from [Tann01] – Copyright Elsevier).

Skeletal maturity indicators are of little use individually because they usually only cover a limited interval of bone development. Some of them are defined as an event (a single stage), the onset of ossification being one example. In order to both span and evenly divide the interval of full development, it is necessary to compound indicators to provide a reliable measure of skeletal maturity. There have been many techniques developed for doing this, although most can be placed into one of two categories: atlas methods and scoring methods.

The atlas methods present the assessor with an atlas of radiographs representing the average skeletal development of healthy children of various chronological ages. The assessor then compares a child's radiograph with the atlas to try to find the radiograph that most closely matches. This matching process is often an overall impression of the radiograph by considering a number of skeletal maturity indicators, although some assessors try to look at individual bones and average the results. The assessment results in the child being assigned a skeletal age that equates to the chronological age of the radiograph the child most closely matched.

Scoring methods score the development stages of individual bones using skeletal maturity indicators. The scores are then combined to generate an equivalent skeletal age that, like the atlas method, is a representation of the child's skeletal maturity with respect to a normal age-related standard. These methods can be as simple as counting ossification centres, through to methods of assessing many stages of many bones. Similarly, the scoring system can be designed for a simple averaging of scores, through to more elaborate systems of

scores that vary with the chronological age of the child [Roch88]. A major advantage of this method is that only one set of radiographs is required for skeletal maturity indicators because the indicators are the same for each sex, and timing differences can be incorporated into the scoring system. Details of some of these methods are presented in Section 1.2.

1.1.2.3 Choice of the hand-wrist for skeletal maturity assessment

The ideal assessment of skeletal age would involve studying many bones of the body. Unfortunately, current costs, long interpretation times, and excessive radiation exposure from radiographs mean that currently this is neither desirable nor practical [Kapl82, p9]. There has been considerable research investigating the portion of the skeleton that best represents the overall development of the skeleton, and numerous methods have been developed for different body regions. These methods range from a detailed evaluation of a single region of the body, through to evaluation of most of the 100 possible ossification centres [Grah72]. Areas that have been looked at include the foot, shoulder, ankle, hip, elbow, cervical spine, and the hand-wrist [Leit87].

As discussed earlier, if a skeletal age assessment method is to be useful, it is important that it be applicable over a wide range of skeletal development. Some skeletal age methods have a limited range of usefulness, some of which is due to the area of the body they examine. For example, the 27-point Sauvegrain scoring system for ossification of the elbow is limited to the period of the pubertal growth spurt when the elbow shows radiographic changes (girls 10–13 years; boys 12–15 years) [Dime05]. The area of the skeleton that has proved most useful is the hand and wrist. The bones of the hand and wrist provide many parallel sequences of changes with physical maturation [Tann01, p1] that cover the period from birth to maturity. Although changes in individual bones tend to stop and start, at any one point in time a number of bones are still undergoing changes and skeletal maturity indicators from each of these bones can be combined to provide a skeletal age estimate. In this thesis the term "bone age" will be used for the assessment of skeletal age using the hand and wrist radiograph.

1.1.3 Applications of bone age assessment to disorders of growth

There are three main applications of bone age assessment in disorders of growth: diagnosis of growth disorders; prediction of adult height; and monitoring of growth hormone replacement therapy.

1.1.3.1 Diagnosis of growth disorders

There are many causes of abnormal growth in children. They can be categorised into hormonal and non-hormonal causes. The non-hormonal causes include genetic disorders,

chronic disease such as kidney disease, and malnutrition. Hormonal causes include growth hormone disorders, sex hormone disorders, diabetes, thyroid disease, and problems with vitamin D metabolism. Many of these causes can be easily diagnosed through clinical symptoms and blood tests. One major exception is growth hormone deficiency. Growth hormone deficiency is especially difficult to diagnose using a blood test because of the pulsatile secretion of growth hormone into the blood stream. A correct diagnosis is important because a deficiency impacts on the growth and development of many organs such as the skeleton and also a number of metabolic functions. A deficiency can also indicate other possible hormone disorders, or a tumour in the pituitary gland in the brain.

Usually investigation of a growth disorder starts with a clinical assessment of a child and measurements such as height, weight, and rate of growth. By combining these with a child's bone age assessment, it is possible to differentiate between broad categories of growth disorders. For example, when there is an intrinsic defect in the skeletal system there is slow bone growth, yet the skeletal maturation does not slow down [Kapl82, p10]. The bone age is either not delayed or the delay is much less than expected based on the height of the child. However, when there are factors involved outside of the skeletal system (extrinsic factors), both growth and skeletal maturation can be retarded, such as in malnutrition where there is impairment of growth and epiphyseal maturation [Kapl82, p10]. A bone age assessment can prove important by identifying discordance between growth and maturation.

Although it appears that bone age assessment has an important role in the diagnosis of growth disorders, what is unclear is the accuracy and reproducibility requirements for the bone age estimate. For example, the Growth Hormone Research Society, in its consensus guidelines in 2000, recommended the routine estimation of bone age for children over the age of one year for those children with growth failure [Soci00]. They also recommended that the bone age assessment should be performed by an"experienced person", although they did not specify acceptable accuracy and reproducibility requirements for bone age assessments in order to gauge what is meant by an"experienced person" There is a distinct lack of consensus, even discussion, in the literature on what constitutes acceptable performance for the assessment of bone age in the diagnosis of growth disorders. When a single measurement of bone age is required for diagnosis, it has been suggested that great accuracy is not required - a tolerance of $\pm$ 0.5 years would be acceptable [Buck83]. Serial observations, designed to show changes over time, demand a much greater accuracy, although the same author did not specify a tolerance for such bone age estimates.

1.1.3.2 Prediction of adult height

An important consideration in the management of growth disorders and short stature is the psychosocial well-being of the child. One aspect of this is the prediction of adult height, often for the purposes of providing reassurance to both child and parents [Mars77, p14]. Most methods of bone age assessment have associated with them some method of predicting final adult height. Modern methods include variables of chronological age, sex, duration of observations of height and bone age, and age in relation to menarche [Buck83]. There is, however, concern about the accuracy requirements for the bone age estimates used in these predictions. In a longitudinal study of 23 boys about to enter puberty, Roemmich *et al.* compared three methods of predicting adult height based on the three methods of bone age assessment [Roem97]. They found that a major source of error in the prediction was the bone age, especially during the pubertal growth spurt. They attributed some of the error in the bone age assessments to the reference populations used to develop the tests, and questioned whether prediction models could be transferred to different population groups. Concern has also been expressed about errors of height prediction models due to errors in bone age assessment [Buck83].

1.1.3.3 Monitoring growth hormone treatment

Many forms of growth failure and growth hormone deficiency (GHD) are treated with synthetic human growth hormone. Growth hormone is expensive and many countries have national programmes for monitoring and regulating it's public use. The estimated annual cost in the United Kingdom of treating a 30 kg child with growth hormone is £6103 if they have growth hormone deficiency, and up to £11132 if they have chronic renal insufficiency [Clin02]. Australia and the United Kingdom treat approximately the same number of children per million population under the age of 20 years, and the Australian cost is estimated at A$16 million per year [Wert96]. Given the costs involved, it is understandable that there is interest in the optimisation of treatment [Wert96].

This optimisation of treatment includes selecting children who are appropriate for treatment (entry criteria); monitoring whether children are responding to treatment (response criteria); only giving as much growth hormone as necessary; and deciding when a child or young adult should stop treatment (exit criteria). Each of the national programmes has different criteria and different emphases on the role that bone age assessment has in treatment optimisation.

In a 1993 survey of growth hormone treatment practices among 251 paediatric endocrinologists, pre-treatment growth velocity was the most important criterion for commencing growth hormone treatment [Wyat95]. Bone age was the only laboratory test always used

by the majority of physicians in deciding when to start and stop treatment. Sixty percent of the endocrinologists used it in their decision to start treatment. Furthermore, bone age was ranked 5th out of 14 criteria both for starting and for stopping treatment.

Serial bone age observations may be used to monitor response to growth hormone. The 2005 Australian guidelines for use of human growth hormone [Heal05] requires submission of a bone age assessment every twelve months. For ongoing treatment the child must have either achieved the 50th percentile of growth velocity for bone age, or achieved and maintained other auxological measures based on chronological age or based on the average parental height. If none of these measures have been met, the growth hormone dose is increased to the maximum allowable for 6 months. If after 6 months none of the criteria have been met, treatment is ceased. This is an example of the indirect use of bone age as part of the monitoring of response and adjustment of treatment.

However, there is controversy over monitoring, with a concern that bone age assessment may be inappropriately used to adjust treatment regimes, and that changes seen may be confused with those seen during puberty [Buck83]. A study using data from the United States National Cooperative Growth Study found that children who had regular bone age monitoring as part of their treatment did slightly better than those who did not [Kauf99]. They concluded that bone age assessment should be part of the follow-up of children treated with growth hormone. However, this study did not look at final heights and because it was a prospective study, there is the possibility of a bias.

An opposing view [Wils99] was based on: a) there are errors in bone age assessment, including erratic longitudinal progression, that could have a substantial clinical impact; b) measurements made during treatment underestimate bone age because there is a delay between cartilage changes and appearance of these changes radiographically (deposition of radiographically visible calcium takes time [Tann01, p47]); and c) a lack of guidelines indicated that clinicians recognise the limited role of routine monitoring. It was unfortunate that this counter argument is based mostly on opinion and qualitative statements. For example, although there is an erratic progression of bone age in normal children, this may not be the case in children who are receiving growth hormone treatment. The role of routine monitoring of bone age remains open to question and it requires further research. A large-scale study of routine monitoring may be possible with the assistance of an automated system for bone age assessment.

1.1.3.4 Other applications

As well as the main applications detailed earlier, bone age has been used in surgical planning and forensic science. In surgical planning, bone age is used in procedures that require

compensation for subsequent growth, such as growth of the mandible [Sato01], and procedures that involve timing of intervention, such as management of conditions of the spine, an example being the correction of sideways bending of the spine [Zeri91].

Bone age is used by forensic scientists mostly for the purpose of providing an expert opinion on age, especially for cases of criminal liability where the date of birth of the child is unknown or cannot be confirmed. It is also one of many tests used for the identification of human remains, especially in mass fatalities [Warr00].

1.2 Bone age and manual methods of assessment

As discussed earlier, 'bone age' is a measure of skeletal maturity based on radiographic assessment of the hand and wrist. There have been many scientific publications looking at various aspects of bone age assessment. Although an extensive literature search has been performed, there is a surprising lack of consensus, even discussion, on what accuracy and reproducibility is required of bone age methods for them to be clinically acceptable. Part of this could be due to the differing roles bone age estimation has depending on the different clinical applications. A literature review shows that there are three main methods of bone age assessment: the Greulich and Pyle (GP) method; the Tanner and Whitehouse (TW) method; and the Fels method (from the Fels Research Institute). The Greulich and Pyle is an atlas method, and the Tanner and Whitehouse and the Fels methods are both scoring methods. Each of the methods is introduced in this section.

1.2.1 The hand-wrist radiograph

The bone age assessment involves an x-ray of the left hand and wrist. The left hand has been chosen by convention because the majority of the population are predominantly right-handed, and with the left hand there is less chance of acquired alterations such as injury [Grah72]. Furthermore, early work on bone age assessment used the left hand probably because the 1906 and 1912 International Agreement for the Unification of Anthropometric Measurements to be made on Living Subjects specified that measurements should be made on the left side of the body and the left extremities [Greu59, p29]. Some investigators recommend x-raying both the left and right hands side-by-side on a single film because of the significant diagnostic value of asymmetry in skeletal maturation [Zeri91]. But studies have shown that any asymmetry is likely to be small and not of practical significance, and it is unnecessary to x-ray both hands [Grah72] [Roch88, p16] [Greu59, p29].

The radiograph resulting from the x-ray is a two dimensional projection of a three dimensional volume of bone, cartilage, and soft tissue. Ignoring scattered radiation, this projection is a summation of the radiographic density of the tissue along the line from the focal

spot of the x-ray tube to the image receptor. The image receptor is usually a light sensitive film in an x-ray phosphor cassette or a cassette with an photo-stimulable phosphor that is read out electronically. The image receptor should provide good spatial resolution and contrast, yet have enough dynamic range to show the tips of the phalanges as well as the wrist.

The left hand should be faced downward, extended, and flat on the x-ray cassette to generate a posterior-anterior radiograph of the hand (Figure 1.4). The axis of the middle finger should be in direct line with the axis of the forearm, and the centre of the x-ray tube above the distal end of the third metacarpal [Tann01, p58]. A tube-cassette distance of 76 cm is recommended by Tanner *et al.* [Tann01, p58], although Roche *et al.* recommend a tube-film distance of 91 cm, on the basis that shorter distances may tend to increase the apparent skeletal age for some methods of assessment [Roch88, p237]. The fingers are spread so they are not quite touching, and the thumb is rotated out to a natural position of around 30 degrees to the first finger. The x-ray beam should be perpendicular to the cassette.

The x-ray is usually performed at a tube potential of 45-60 kVp. If the x-ray exposure is too low, as can easily occur with incorrect use of photo-stimulable phosphor plates, then image quality is compromised because of increased noise. An example of the impact of this noise is an inability to determine the curvature of the corners of the epiphyses, which is an important feature in some skeletal maturity indicators.

Correct positioning is important because poor positioning can change the appearance of some bones. For example, if the projection of the x-ray beam is not perpendicular to a joint, and therefore not tangential to the joint surface and physis, then the gap between the epiphysis and metaphysis can appear artefactually narrow. Similarly, obliquity can cause a change in the ratio of the width of an epiphysis to the width of a metaphysis. Both of these features are used in some skeletal maturity indicators.

1.2.2 Skeletal maturity indicators of the hand and wrist

Each method of bone age assessment tends to use different skeletal maturity indicators. However, there are some common elements to these indicators. As discussed in Section 1.1.2.2, the onset of ossification of bones or epiphyses and the fusion of the epiphyses are clear skeletal maturity indicators. The onset of ossification is usually the first step on the skeletal maturity scale. The first carpal bones to appear are usually the capitate and hamate. Ossification of the phalangeal epiphyses usually proceeds from the first to the fifth finger, and usually from the distal to the proximal phalanges [Leit87]. As the bones and epiphyses develop, the radiographic outlines of the bones change shape, some of which is due to their interaction with the articulating surfaces of nearby bones.

Figure 1.4 X-raying the left hand and wrist for a bone age assessment.

Figure 1.5 shows three radiographs that demonstrate some of the bone changes that occur with skeletal development. As the child develops, the carpal bones appear at various stages then change shape as they develop articulating surfaces. The epiphyses of the long bones appear, change in shape, and then fuse with the long bones.

Some of the maturity indicator grades are based on metric procedures, although absolute sizes are inappropriate for assessment of maturity and it is usually the relative sizes of bones that are used [Roch88, p49]. For example, in the assessment of epiphyses, the width of the epiphysis is compared with the width of the metaphysis.

There is debate in the literature about the variability of individual maturity indicators. According to Roche *et al.*, the sequence of skeletal maturity indicator grades is approximately fixed for any one bone, although within the whole hand it is "far from fixed" [Roch88, p22]. A common example of this is the variation shown between the carpal bones and rest of the bones in the hand and wrist. This view is not shared by Tanner *et al.*, who claim that their stages and sequences are the same in all populations, and are even unaffected by starvation [Tann01, p18].

It is generally accepted, however, that the skeletal maturity indicators for any one bone differ in significance as the bone matures. If a bone lasts in a particular developmental

state for a long period of time then it is less useful in the skeletal maturity assessment than a bone that passes through a number of states quickly [Tann01, p6]. The Fels method is one system that makes use of this idea [Roch88].

1.2.3 Methods of bone age assessment

The following summarises the three main methods of manual bone age assessment based on hand-wrist radiographs.[1] The differences between the methods are discussed in Section 1.2.3.4

1.2.3.1 The Greulich and Pyle method

The Greulich and Pyle method (GP method) is an atlas-based method [Greu59]. The assessment is performed by comparing a child's hand-wrist radiograph with an atlas containing standard radiographs for a range of skeletal ages. The comparison starts by finding which standard superficially matches the child's radiograph. Greulich and Pyle suggest using the child's chronological age as an estimate of the skeletal age, but then comparing the child's radiograph with standards before and after the chronological estimate. A preliminary selection is then made by comparing the child's radiograph with the standards and looking at gross maturity indicators for the child's period of skeletal development, such as an impression of the degree of epiphyseal fusion. The method then requires the assessor to work systematically through the chosen radiograph, comparing each of 31 bones and sesamoids in the hand-wrist with the standards in the atlas (Figure 1.6). An age is assigned to each bone using data from tables associated with the standard that contains the closest match to the bone. If no match is found for a bone, the age is estimated from the closest matching radiographs. To assist the assessor in the matching process, the atlas contains outline drawings and text for a range of skeletal maturity indicators. Each radiograph also has a written description of important skeletal maturity indicators for the associated skeletal age. If the age of each bone corresponds to a single standard in the atlas then the child's bone age is that of the standard. If there is no exact match and the results are intermediate between two standard radiographs, then an intermediate estimate between the two corresponding ages is used. Greulich and Pyle provide no recommendations on how to perform this calculation, but in implementing the Greulich and Pyle method, one experienced assessor used the median of the ages of the individual bones [Roem97].

The atlas consists of 27 standard radiographs for females and 31 standard radiographs for males. The standards start at intervals of 3 months, then go to 6 months at 1.5 years, but

[1] There are other methods of bone age assessment. For example, Eklöf and Ringertz introduced a method that used 10 absolute measurements of bone size and a two-sigma statistical variation limit for each measurement [Eklo67]. This method has been compared with the GP and TW3 methods, but was shown to overestimate bone age in young children, and underestimate bone age in older children [Hait06].

(a) Radiograph of a 3 year old child showing first appearance of the carpal bones and early appearance of the epiphyses of the long bones.

(b) Radiograph of a 10 year old child showing the advanced development of the carpal bones and metcarpals and phalanges.

(c) Radiograph of a 16 year old child showing overlapping of the carpal bones and fusion of the metacarpal and phalangeal epiphyses.

Figure 1.5 The hand-wrist radiographs of three children, demonstrating variation in the bones with age.

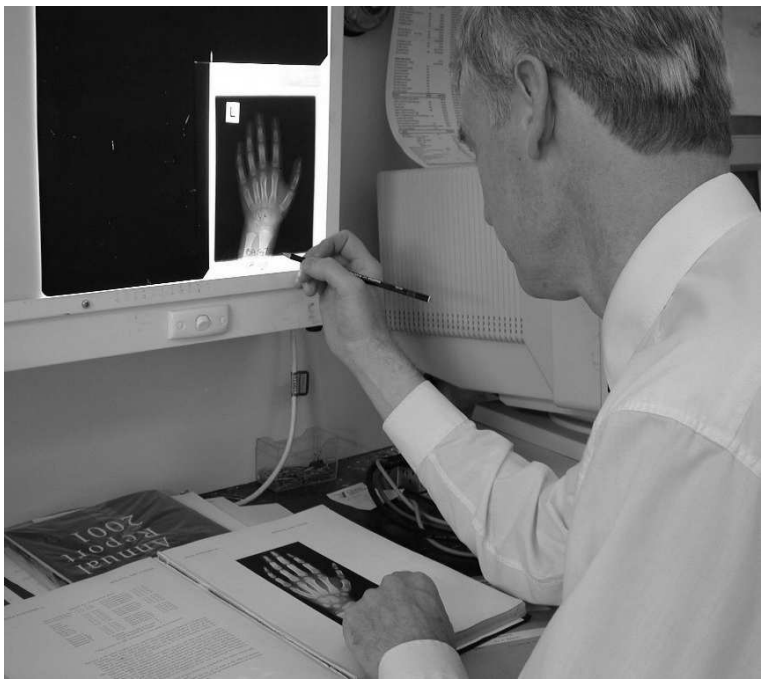


Figure 1.6 A radiologist reporting a bone age radiograph using the atlas-based Greulich and Pyle method

reduce to annual increments from the ages 5 years to 18 years in females and 19 years in males, with the exception of an additional standard at 13.5 years in females and 15.5 years in males because of the relatively rapid rate of skeletal maturation during puberty. The lack of additional standards around the time of the pubertal growth spurt has been seen by some as a major disadvantage of the GP method because it reduces the accuracy of the bone age assessment at this important stage of development, and results in increased inter-observer error [Dime05]

Each standard used in the preparation of the GP atlas was chosen from one hundred children of the same age and sex. The radiographs of these children were ordered by their relative skeletal status and one was chosen that "was most representative of the central tendency, or anatomical mode" [Greu59, p32]. The one hundred radiographs used for each standard were drawn from radiographs of 1000 children involved in a long-term human growth and development research study run by Professor T Wingate Todd (Western Reserve University of Medicine) in the Cleveland, Ohio, region between 1931 and 1942. This study only admitted children referred by a paediatrician. This lead to a referral bias of children of above average socioeconomic and educational status, and "all the children were white, all had been born in the United States, and almost all were of North European ancestry" [Greu59, pxii]. This is a serious criticism of the GP method because some investigators believe that the sample population of upper-middle-class children matured rapidly

and this meant that until recently the standard radiographs in the atlas were considered advanced for chronological age [Tann01, p17] [Cox97].

1.2.3.2 The Tanner and Whitehouse method

The Tanner and Whitehouse (TW) method is a score-based method [Tann01]. The method involves independent staging of up to 20 bones of the hand and wrist. The bones are divided into 8–9 discrete stages of development that span from the onset of ossification to full maturity of the bone. The stages were originally chosen by trial so that the step from one stage to another was not so large as to lose information, and not so small as to "cause confusion and rating errors" [Tann01, p5]. The stages are assigned a letter from A to H, with an A meaning that no sign of the bone was present. There are up to three written criteria for each stage and a diagram of typical appearances. If one criterion is specified, it must be met for the bone to be assigned that stage. Only one criterion needs to be met if there are two listed. However, if there are three criteria listed, then two criteria need to be met. In addition, for any stage to apply, the first criterion of the previous stage must have been met [Tann01, p60]. Some of the criteria include relative distances between bones and the ratio of bone sizes, so an instrument for measuring distances is required. No overall standard radiographs are present in the book on the method, only cropped radiographs that are typical for each bone stage being considered. These are only meant to be a guide, but many users turn to them when they find the written criteria difficult to understand [Cox96]. The stages and criteria are the same for both females and males, and are independent of ethnicity and socioeconomic status [Tann01, p18].

Associated with each stage is a maturity score that was calculated to minimise the overall disagreement between bones of an individual [Tann01, p3, p105]. These scores are gender specific and were developed by analysis of a large sample of bone ages from infancy through to adulthood. The scores were scaled so that the sum of the scores was in the range 0 (invisible) to 1000 (full maturity), representing the span of development. To complete the TW assessment one finds the stage of each bone being considered, notes the score for the stage, then sums the scores across all bones to produce a skeletal maturity score.

The skeletal maturity score is a bone maturity measurement that Tanner and Whitehouse propose is "independent of such influences as secular growth changes, socioeconomic class and ethnic group" [Tann01, p9]. Part of the design of the TW method was to produce a measure of maturity that was independent of age. They consider the skeletal maturity score to be a measure that is comparable to other body measurements such as height and weight. Many clinicians, however, prefer the use of bone age. Percentile plots of skeletal maturity score versus chronological age were developed in order to convert a child's skeletal maturity score into a bone age. The bone age is the chronological age correspond-

ing to the 50th centile of the skeletal maturity score. This percentile data is population specific [Tann01, p10], whereas the skeletal maturity score is not.

The Tanner and Whitehouse method was originally introduced in the 1960s and was based on a cross-sectional sample of some 2700 healthy, mostly working-class, Scottish and English children of the 1950s and 1960s [Tann83, p14]. The original method involved staging of 20 bones: the 13 long bones - radius; ulna; metacarpals 1, 3, and 5; proximal phalanx 1, 3, and 5; middle phalanx 3 and 5; distal phalanx 1,3, and 5; and the 7 carpal bones – capitate, hamate, triquetral, lunate, scaphoid, trapezium and trapezoid. This method was revised and superseded in 1975 and 1983 (second edition) by the TW2 method [Tann83]. The TW2 method had three scoring systems based on different sets of bones: TW2 20-bone (as above); TW2 RUS – the 13 bones Radius, Ulna and Short bone score (finger bones); TW2 Carpal - the 7 carpal bone score. The separation between TW2 RUS and TW2 Carpal was based on the concern that carpal bones gave different maturity information than the long bones, and Tanner *et al.* thought that it was important to separate them [Tann83, p4]. With the TW2 Carpal scoring system each carpal bone was given an equal weight compared with the other carpal bones. With the TW2 RUS scoring system, the metacarpals and phalanx bones provided essentially the same maturity information, so they were given a lower weight; the radius and ulna were given higher weights. These weights were built into the score for each stage, so to apply the scoring system it was simply a matter of summing the scores for the bones being assessed.

Because of the weights used in the TW2 20-bone scoring system, it was essentially an average of the TW2 RUS and TW2 Carpal scores. It was not surprising, therefore, that this scoring system was dropped when the TW3 method was introduced in 2001 [Tann01]. The main change with the TW3 method was the introduction of new reference populations for conversion of the TW2 RUS scores to bone age. The Carpal scoring system was unchanged.

1.2.3.3 The Fels method

The Fels method is similar to the TW method in that it is a score-based method [Roch88]. It was developed in the 1980s and is based on 98 skeletal maturity indicators and 13 measurements. These indicators and measurements were arrived at by a complex series of steps that involved using already published indicators and rewriting their descriptions to make them clearer, then going through an extensive task of reviewing, grading, and testing the indicators [Roch88, p44]. For any given chronological age, only a quarter of the indicators in the Fels method are used.

The assessment procedure starts by using the chronological age and sex to find which indicators should be assessed for any given bone (this information is provided in a table). A

total of 22 bones are assessed: the same bones as the TW2 20-bones system, plus the adductor sesamoid (an ossification centre near the end of the first metacarpal), and the pisiform (an ossification centre near the triquetral bone). The applicable maturity indicators for each bone are then graded using a combination of written descriptions, line drawings and sample radiographs. The indicators can be binary (only two grades), more than two grades, and continuous. Respective examples of these are the onset of ossification, the shape of bones, and the ratio of epiphyseal to diaphyseal widths. Once the bones have been graded for each indicator, the grades are entered into a computer program that produces a bone age estimate, along with a standard error for the estimate. The computer program uses a probabilistic model of maturity that is calibrated for each indicator. The bone age is estimated from the "maximum value of the product of trace lines for the observed data" [Roch88, p311].

The reference population for the Fels method was 667 children enrolled in the Fels Longitudinal Study of Growth and Development. These children lived in southwestern Ohio and were divided between cities, small towns, and farms. They had a socioeconomic distribution that was similar to national samples, but with an under representation of the lowest socioeconomic group [Roch88, p41]. Serial radiographs of children between 1932 and 1977 generated 13823 radiographs. These radiographs produced good coverage of the span of development from 1 month to 22 years. The early radiographs were at ages of 1, 3, 6, and 9 months, then every 6 months from 1 year of age until 18 years, and finishing with x-rays at 20 and 22 years.

Although the TW and Fels methods may appear similar, there are significant differences. The first difference is that the TW method requires no knowledge of chronological age, whereas chronological age is necessary for the Fels method. Interestingly, one group recommends that if the Fels estimated bone age differs from the child's chronological age by more than two years, then the assessment should be repeated using the extra indicators that would be required if the estimated bone age was used as the chronological age [Roch88, p238]. Knowledge of chronological age in the GP method has been shown to not influence the accuracy of bone age estimates, although assessors may be biased towards interpreting the bone age as normal if chronological age is known before the assessment [Bers01]. Given the objective nature of scoring techniques used in both the TW and Fels methods, knowledge of chronological age is unlikely to be a problem.

The second difference is the scoring of the bones. The TW method uses maturity indicators to assign a stage to a bone. The Fels method grades each applicable maturity indicator for the bone, and, therefore, produces multiple grades for each bone, not a single stage. Similarly, some relative measurements are used to stage bones in the TW method, whereas the Fels method uses these relative measurements as continuous maturity indicators. These

are significant differences in the approach to skeletal maturity estimation.

1.2.3.4 Accuracy and differences between the three systems

The main difficulty with assessing the accuracy of any bone age assessment system is the lack of a gold standard. The best that can be achieved is the assessment of bone age in a large group of normal, healthy children at a single point in time. The bone ages are then compared with the chronological ages of the children, under the assumption that across a large group of children the mean differences between their bone age and chronological age will be zero. This is a cross-sectional comparison because it captures children at a single point in time. This method has been used in many studies comparing the different methods of bone age assessment, although comparative studies also tend to compare bone age measurements against each other, not just against chronological age. When chronological age is not used the accuracy cannot be established, only the consistency between methods.

The accuracy of the methods of bone age assessment depends on the reference population used to develop them, and its relationship to the patient population being measured. It is also dependent on the implementation of the techniques of assessment. If someone is poorly trained in a method, then it is likely that their method of assessment will not match that which was used to analyse the reference population. Hence, their bone age results will be inaccurate. Part of the problem with the accuracy arises from the subjectivity in interpreting the skeletal maturity indicators of the radiograph. All manual methods suffer from this, but the TW and Fels methods are thought to be more objective than the GP method because the bones are considered in a systematic way, and in isolation. This is thought to be one of the reasons why the TW method is slightly more reliable than the GP method [Roch88, p36].

When compared with each other, the three methods produce different results. Some of this relates to the differences in reference populations used to develop the methods, some is due to subjectivity, and some is probably due to the skeletal maturity indicators chosen for an assessment. The problem of different reference populations explains many of the research findings of differences between the methods. For example, in a study looking at the prediction of adult height for 23 boys during puberty, bone ages were assessed using the GP, TW2, and Fels [Roem97]. The unique aspect of this assessment was that it was performed by three experts in the area, including Tanner (TW2) and Roche (Fels), and the GP method used a bone-by-bone assessment. Although the sample size was small, this was a longitudinal study following the boys for at least 6 years and with bone age assessments every 8 months. The mean bone age results showed that for boys in the age range 9-15 years, the TW2 bone ages were consistently more advanced than the GP and Fels results.

A comparison of the Fels method with both the GP and TW2 methods was performed using patients from the Fels longitudinal study [Roch88, p265]. On average the Fels bone ages were larger than the GP bone ages in younger children, but in young teenagers the Fels bone ages were smaller than the GP bone ages for both sexes. The opposite was found when comparing the Fels and TW2 methods. In the very young the two methods gave comparable results, but by the age range 6.5 – 9 years the TW2 results were up to 2 years more advanced than Fels bone ages in girls, and 1.7 years more advanced in boys. In both sexes the differences between the two methods decreased as the children approached puberty. Given the similarities between the Fels and TW2 methods, the most likely explanation is the influence of different reference populations.

In addition to the reference populations, there are other inherent differences between the methods. One of these is the impact of variability in the maturity of different bones within a single radiograph. This is a major limitation of the GP method [De L99], but the TW and Fels methods partly overcome the problem because they consider each bone in isolation and then combine scores at the end. With the GP method, each bone is meant to be considered in turn, but as mentioned in Section 1.2.3, there is no formal method of combining them. The result is that when there is dissociation within the radiograph such as advanced carpals compared with the long bones, it can be difficult to assign a bone age using the GP method.

There remains no agreement on which method of bone age assessment is the best to use, although it has been suggested that "some are better in specific disease states" [Alba95]. In one survey 182 paediatricians in England and Wales indicated that 76% of bone age assessments were performed using the GP method, 20% used the TW method and 4% used other methods [Buck83]. The GP method is most widely in the United States [Zeri91], although the TW method is preferred in Europe [Gree01, p178]. Some assessors consider the TW2 method too laborious for routine use. Yet the GP method requires just as much care to achieve accurate bone age estimates if it is performed as per the GP atlas [Mars77]. Part of the popularity of the GP method is the ease with which the method is performed, and the short time it takes to complete an assessment. This is because many institutions and individuals perform a method based on an overall impression of a child's radiograph in comparison with the standard radiographs [Bull99]. This is essentially the method used for the preliminary selection of a standard and is an abbreviated GP method. This simplifies the assessment and reduces the time it takes to complete. Typical times for the GP method are 1.4 minutes, and 7.9 minutes for the TW2 method [King94].

Lastly, there is a difference in the maximum bone ages for the three methods of assessment. They are 16.5 years for the TW3 method [Tann01], 18 years for the GP method [Greu59], and 18 years for the Fels method [Roch88]. The significance of these differences is not

known.

1.2.3.5 Intraobserver and interobserver variability

Intraobserver variability, or reliability, refers to the stability of an individual's assessment of bone age between two points in time. Interobserver variability is the agreement between two or more individuals performing an assessment. Both of these are important parameters for any bone age assessment method because they lead to a reduction in accuracy and reproducibility of bone age results.

A review of published studies on variability in the GP method, showed that the interobserver variability for the GP method ranged from 0.10 years to 1.04 years [Bers01], and although the intraobserver variability ranged from 0.09 years to 1.19 years, overall they found that the intraobserver variability was less than the interobserver variability. A study comparing the GP and TW2 methods tried to simulate the clinical practice of an inexperienced observer by using three, second year radiology registrars [King94], and found a TW2 intraobserver variation of +1.5 to -0.9 years for a 95% confidence interval. This variation is considerably wider than $\pm$ 0.5 years for experienced assessors [Tann83, p14]. Although the interobserver differences between the GP and TW2 method showed no statistically significant difference, an inexperienced observer is likely to have a poor intraobserver variability for the TW method [King94] (unfortunately, this study did not look at intraobserver variability for the GP method.).

In a study of two trained technicians using the TW2 method [Zach83], the technicians achieved interobserver and intraobserver errors similar to findings by other researchers. However, they found two cases with the TW2 method in which one of the technicians estimated bone ages that were more than 1 year different from three other observers (two of whom were experienced assessors). On re-evaluation of the two cases, the carpal and phalangeal bones were found to be significantly different in their levels of maturity. It required a more experienced assessor to compensate for these differences. A similar phenomena has been observed when using the GP method [Roch70a]; when a radiograph was assessed for a 'whole-hand' bone age, the correspondence between paired radiographs was lower if there were marked maturity differences between the bones.

The unresolved question with intra and interobserver variability is "what performance is considered acceptable?". As discussed in Section 1.1.3, the acceptable accuracy and reproducibility are likely to depend on the clinical application. For an initial diagnosis, an accuracy of $\pm$0.5 years has been suggested, but serial observations demand greater precision [Buck83]. The above discussion shows that this could only be achieved by experienced assessors, and even for them the intraobserver and interobserver variability is larger than

desirable.

One approach to the problem of interobserver variability is to use the same observer for all bone age assessments. In a survey of United Kingdom bone age assessments, approximately two thirds of the assessments were performed by three or more individuals in a institution [Buck83]. Apart from the scientific evidence for intraobserver and interobserver variability, there is some strong feeling in the literature about misuse of methods of bone age assessment. This is best summarised by Graham's statement "valid skeletal age assessment presupposes a working knowledge of the fundamental concepts and tools involved, and an amateur interpretation is often worse than none at all." [Grah72].

1.3 Justification of automated assessment of bone Age

1.3.1 The reduction of observer variability through automation

For bone age assessment to be of use in both clinical and research environments, it is important that the bone age result be accurate, the results be reproducible between assessors (observers), and that the results be consistent with serial measurements of bone age. Assuming an automated bone age system can reasonably duplicate the tasks of a human assessor, then subjectivity of the assessment would be eliminated as a possible cause of accuracy loss. The implementation of a hybrid system combining a range of skeletal maturity indicators from the different methods should address accuracy problems due to inherent limitations in the bone age assessment method. Finally, having an automated system highlight problems with the dissociation of bone maturity within a single radiograph would provide an opportunity for an expert assessor to intervene in the assessment.

The findings of large intra- and interobserver variability and the unresolved questions of what variability is considered acceptable are motivation for development of an automated system for bone age assessment. If such a system does not involve random processes, such as Monte Carlo methods, and the algorithms are stable and tolerant of small changes in radiographic features, then there is no reason to expect an automated system to produce variability between bone age assessments. This should be the case for both reassessment of the same radiograph, and assessment of pairs of repeat radiographs at the same age. The objective of the automated system would be to eliminate observer variability, to leave human maturational variance as the main contributor to uncertainty in bone age assessment.

1.3.2 The feasibility of using novel measures of bone age

The significant differences in skeletal maturity indicators for the TW and Fels methods, and the differences in how the skeletal maturity information is combined, suggests the possi-

bility of other, novel measures of bone age assessment that have not yet be developed. For example, Roche *et al.* found through their extensive analysis of skeletal maturity indicators that there were 13 indicators in boys that were not useful when applied to girls [Roch88, p57]. They decided to drop these 13 indicators from the Fels method and this resulted in 98 indicators that could be used for both girls and boys. However, it is important to note that their reason for looking at the indicators separately for boys and girls was they thought that some of the indicators might have been influenced by differences in muscular development specific to boys [Roch88, p48]. Although they did find differences between girls and boys, they chose not to include them, probably in the interests of simplifying the Fels method. This is an example of where automated analysis would allow extra indicators to be incorporated into the system, and could even allow indicators to be added that were difficult for a human assessor to determine (for example, quantitative measures like ratios of areas). An automated system could easily cope with the complexity overheads of extra indicators as long as computational overheads did not become prohibitive. Ideally, an automated system should have the flexibility to allow new measures of bone age to be added. Existing, manual methods are much more difficult to modify, mostly because of the need to retrain the person performing the assessment. Furthermore, there is the possibility of measures that a computer is good at quantifying but not a human; measures such as area and contour angularity.

As mentioned in Section 1.3.1, one of the concerns about bone age assessment is when there is a lack of maturity consistency between individual bones [Tara76]. An automated system could provide data management facilities that would make it possible to highlight bones that show dissociation with other bones in the hand and wrist. Furthermore, a system could track both individual bones and groups of bones across serial investigations. In this respect, differences between bones could be useful, rather than simply trying to assign a single bone age to the whole radiograph.

In addition to new measures of skeletal maturity, an automated system could allow changes to the methods by which bone ages are assigned to skeletal maturity scores or other measures. For example, the biological weights used by Tanner *et al.* in the development of the TW method did not take into account the duration of the stages of the bones [Tann01, p6]. They thought that a bone should have a lower weight if it lasts in a stage for a long time. If a bone passes through a stage quickly, then it should have a higher weight. This concept was used in the Fels method [De L99], but Tanner *et al.* felt that a system that included this extra component of variable bone weights would be troublesome to use. This type of change to a bone age method would be much simpler to incorporate into an automated system.

1.3.3 Population specific datasets

There are two aspects to the need for population specific datasets. Firstly, much of the data used in older bone age assessment methods was based on historical series of radiographs for particular populations [Cox97]. There has been a worldwide trend of children maturing earlier and a call for new reference standards has being made [Grah72, Cox97]. Secondly, some investigators believe that the use of a particular bone age method should be restricted to children who share the same socioeconomic and genetic characteristics as the reference population from which the method was developed. Both of these aspects of the reference populations have been addressed in research, yet many studies appear to raise more questions than provide answers. For example, a literature review of the effects of ethnicity on skeletal maturation found that there was little influence of ethnicity on bone age assessments, but there was a socioeconomic effect [Schm00]. In a study of the applicability of the reference population for the GP method, different ethnic groups – Asian, Hispanic, Black, and White - had statistically significant differences in bone age and chronological age across all groups, except Asian girls [Onte96]. Other population-specific differences have been demonstrated by another group who found that in 1986 Japanese children were maturing earlier than children in Belgium, United Kingdom, and South China when assessed using the TW2 RUS skeletal maturity score [Mura97]. These examples are typical of the conflicting findings in the published literature.

One approach to the problem of difference in reference populations would be to use a single method of skeletal maturity assessment across a number of different populations of children, and to look at a measure of skeletal maturity that is independent of any one reference population. Tanner *et al.* proposed that the skeletal maturity score from the TW2 (TW3) method was a measure of skeletal maturity that was independent of the reference population [Tann83, p9]. This has been backed-up by other investigators who believe that because skeletal maturation is orderly across populations, even in those with malnutrition, a skeletal maturity score may be a more appropriate measure of skeletal maturity than bone age [Roem97]. Instead of trying to relate skeletal maturity to equivalent development in normal children of a specified chronological age, the skeletal maturity scores are used directly. The basis of this is that the score only relates to radiographic appearance of the bones and instead of determining a score - then an equivalent bone age and working out if the bone age is normal - normal ranges are calculated for skeletal maturity scores directly. Hence, two children with the same score should be at the same level of skeletal maturity, independent of the population from which they are drawn.

Establishing normal ranges for skeletal maturity scores, or processing new reference populations for existing methods would require a lot of work. Although there would be much effort required in recruitment, clinical measurements and demographics collection (includ-

ing the challenging tasks of socioeconomic and ethnicity assessment), a significant amount of work would be required for the bone age assessments. This task could be simplified by an automated bone age system that would allow for the processing of large numbers of radiographs, without the need for duplication of assessments ('second reading' of the radiographs).

1.3.4 Development of a research tool

Use of an automated system to improve the accuracy and precision of bone age measurement may help with investigation of questions surrounding the use of the assessment for monitoring skeletal maturity during growth hormone replacement therapy. It is expected that an automated system would reduce the costs of bone age assessment through a reduction in time that radiologists would need to report on bone age radiographs.

Part of the problem with the interpretation of research studies involving the use of bone age is the uncertainty in the results. Using an automated system to reduce uncertainties due to the bone age assessment may improve the power of the study and reduce the number of participants required to achieve a significant result.

Using currently available x-ray tube and x-ray generator technologies, combined with digital x-ray capture devices, it should be possible to construct a very low radiation dose, portable, automated bone age assessment system for screening of children. Roche *et al.* state that "assessment of skeletal maturity is an important part of any epidemiological study involving the physical status or performance of children" [Roch88, p10]. A valid and reliable automated bone age assessment system could be an important tool in many research studies.

1.4 Conclusions

Maturation is a difficult process to define and quantify, yet it is important in the understanding and management of growth disorders. Although it is by no means perfect, there is a well-established relationship between biological maturation and skeletal maturation. Bone age is one measure of the amount of skeletal maturation that a child has undergone - their level of skeletal maturity. It is based on an assessment of skeletal maturity indicators in a radiograph of a child's left hand and wrist.

This chapter has introduced background material required for the thesis, including a discussion of clinical applications of bone age assessment, highlighting the lack of information about accuracy and reproducibility requirements for bone age assessment. Although these requirements are likely to depend on the clinical application, studying them requires a

reliable research tool for bone age estimation. In order to investigate application-specific tolerances required for a test, this argument says that the test should be both accurate and reproducible (precise) to start with.

The thesis for this research is that it is possible to develop an automated system that will produce an accurate and reproducible measurement of bone age. Such a system would be valuable in both clinical and research practice. Chapter 2 reviews work published on this topic by other investigators. It then introduces the research components of this thesis.

Chapter 2

Computerised assessment of bone age

It is technically feasible to automatically assess bone age deriving skeletal maturity information from radiographs of the hand and wrist. However, this task is not trivial. The hand-wrist radiographs contain a complex collection of bones that change shape and size over time, with some bones overlapping, and others not even being present at various stages of development. As demonstrated by the bone age variability discussed in Chapter 1, analysing the bones and deriving a bone age is a complex task, even for humans. The image processing task required to analyse the radiograph is equally as complex. Image noise and the poor contrast of some bone edges further add to the challenge of automated assessment. There is certainly no guarantee of achieving comparable or superior results to the manual methods.

Researchers have recognised the importance of automating the assessment of bone age. Much of this research has focussed on specific components of the automation, such as initial segmentation of the radiographs. But some research has resulted in bone age assessment systems that have been used in clinical research. The first section of this chapter reviews and summarises the research on bone age assessment systems. A model of a system for the computerised assessment of bone age is then introduced. The main components of this model are then described, with details reserved for other chapters of this thesis. This is followed by discussion on the role of the carpal bone analysis and the potential for improvement of automated bone age assessment by including this analysis in the bone age assessment of older children.

2.1 A review of computerised assessment of bone age

An ideal system for the automated assessment of bone age should start with a digital radiograph of the left hand and wrist, either from computed or direct radiograph, or from

digitisation of an x-ray film. The system should be tolerant of a range of exposure parameters used in taking the radiograph, tolerant of noise and artefacts, and able to accommodate a significant amount of poor radiographic technique (for example, poor collimation of the x-ray field or poor positioning of the hand). The system should produce a bone age based on a specific reference population, or produce a measure of skeletal maturity that can be converted to a bone age for a specified reference population (as with the TW method). Although the result would ideally be more accurate and reproducible than a radiologist, the bone age estimate should also include a measure of uncertainty. The system should store all results for a child, and allow for ongoing monitoring of skeletal maturity for the purposes of therapeutic intervention. Finally, the system should be easy to adapt to a new reference population.

A review of the literature, commercial products, and patents showed that at this stage no such system exists [1]. Many research groups have investigated aspects of such a system using computerised methods. This research can be divided into three broad categories: computer-assisted manual assessment, user-assisted computerised assessment, and fully automated assessment. The intention of this section is to demonstrate these three categories through a review of different approaches. It is important to note that research on computerised bone age assessment is a continuum from data-keeping for manual methods, through to fully automatic systems that integrate into clinical environments. The division into three categories is only designed to assist discussion.

Within all of the three categories the tendency has been to approach the bone age assessment as separate tasks for each region of interest of the hand-wrist radiograph, as is recommended by the TW and Fels methods. The bone age is then derived by combining information from each region. These regions of interest (ROI) are the phalanges (PROI), the carpus (CROI), and the radius-ulna (RUROI). The skeletal maturity indicators are different for each of these regions, and the computerised methods need to be correspondingly different.

Although the following discussion works towards fully automated systems, in many respects such bone age assessment can only ever be considered computer-assisted because a human, often a radiologist, takes overall responsibility for the final reported bone age estimate [Piet95] [Gros95]. To assist with this requirement, Gross *et al.* recommend that the computerised bone age result be used to select a comparison image from the GP Atlas so that the human user can perform a simple visual check of the result [Gros95]. A more sophisticated method of automatically selecting a reference image from a database was implemented by Pietka *et al.* [Piet05] [Piet03b].

[1] A recently available commercial product comes close to meeting many of these requirements (see Section 2.1.3.9)

2.1.1 Computer-assisted manual assessment

When performed according to methods specified in the individual atlases, the GP, TW, and Fels methods are all complex methods of bone age assessment (Chapter 1). It is not surprising that some research has focussed on removing some of the laborious processes of the assessment and allowing the person performing the assessment to concentrate on the task of assessing skeletal maturity indicators. The simplest approach is the use of a computer to record the observations of the assessment. This has mostly been performed for the TW method because skeletal maturity stages are discrete and the calculation of a skeletal maturity score, and corresponding bone age, can be easily performed using computer software. For example, the TW3 book by Tanner *et al.* includes software for calculating a child's bone age by entry of the stages of each of 13 bones of the TW-RUS method [Tann01]. Similarly, the Fels method requires computation using computer software to calculate a bone age estimate [Roch88].

More complex computer assisted methods have been developed that assist a user by presenting them with comparable atlas images and descriptions of the skeletal maturity criteria for the TW method. One system included a database for managing both patient results and the data used in the bone age analysis [Niem02]. In a comparison between bone age assessments performed using this system and those performed manually, the interobserver variability was comparable to that of the manual TW2 method. The computerised method simplified some parts of the assessment and had good data management, but even so, this had little impact on the variability of bone age estimates.

2.1.2 User-assisted computerised assessment

User-assisted computerised assessment refers to when a user is required to add input to assist the software. This approach recognises that there are aspects of bone age assessment that may require the skill and judgement of a human observer to achieve good results. Such assistance could be required because of the cost and difficulty in automating certain aspects of the assessment, or if current automatic methods produce unreliable results compared with assistance from a user.

2.1.2.1 The CASAS system and use of a continuous scale: Tanner *et al.*

Early in the development of the TW method, Tanner and Cameron proposed that the process of allocating TW bone age stages was something that was suitable for a computer to perform [Tann01]. With the assistance of Tanner, Discerning Systems Inc. developed a computer-aided skeletal age scoring system (CASAS) based on the TW2 RUS method. With user-assistance, this system digitised the x-ray bone by bone using a light box and monochrome video camera. The system analysed the digitised information using methods

of classification statistics. (this system was later replaced with a digital version that was designed to process digital radiographs, although the processing remained essentially the same). Each bone was positioned and zoomed on the camera with the aid of an overlay template. If required, the captured image was histogram equalised and filtered to remove "extraneous details and radiographic imperfections" [Tann94b].

The computer performed a template matching assessment using a two dimensional fast Fourier transform. The bone was matched by finding the best average template that minimised the root-mean-square error between the coefficients of the Fourier transform of the bone, and the coefficients of the Fourier transform of the bone template [Tann94b]. This produced a discrete stage for the bone maturity stage, but Tanner *et al.* took it a step further and used the root-mean-square error for the two stages above and the two stages below the best match. They fitted a Gaussian function to the resulting five root-mean-square error values, and used the mean from this fit as the bone maturity score. This produced a continuous bone stage between 0 and 9.0. The stage for each bone was used to calculate an overall estimate of the bone age using the TW2 system. Personal communication and internet searches indicate that this product is no longer supplied by Discerning Systems Inc. (Jan 2009).

An important contribution to the performance of this system was how representative the bone templates were of the TW bone maturity stages. The templates were generated by averaging the Fourier transform coefficients from at least 10 radiographs of each bone stage. For some of the middle stages, as many as 30 radiographs were available and were used to generate the template. The radiographs were from the TW 'Basic Series', a series of radiographs from a London longitudinal study in which bone age had been assessed and reassessed over many years [Tann94b]. Although the radiographs from this series were used in the development of the standard for the TW skeletal maturity, they were not used in the development of the actual bone scoring system, according to information from Tanner *et al.* [Tann83, p4], an approach that could be flawed. The TW method places a large emphasis upon the verbal criteria that "describe shape and density markings of each bone" [Tann83, p4]. Given that the Fourier transform is linear, the process of Fourier transforming the images and taking the average of the coefficients is the same as taking the Fourier transform of the average image. Minor amounts of misregistration and scaling would result in a reduction in the high spatial frequency components in the average image (they would be 'blurred' by the averaging). It is difficult to say how the coefficients of the Fourier transform provide a measure of the shape and densities of the bones, to meet the verbal criteria of the TW method. If the method is based on the ability to match the examined radiograph with the template image, then it is unlikely that it is doing so on the basis of morphology. The template images are critical, and the choice of the source images to create these template images becomes important.

Of all the automated bone age assessment methods to date, the CASAS system has been tested the most. This testing included evaluation of the system against radiographs used in the development of the TW method, as well as radiographs of normal children and those with proven pathologic conditions. A number of studies have investigated differences between the CASAS and manual TW bone age assessments, including how often intervention by the user was required, how well the manual and CASAS methods compared in calculating individual bone stages, and the overall bone age results (refer to Appendix A). They showed that on average one manual insertion was required per radiograph, the repeatability of the CASAS system is better than the manual TW method, and there is reasonable agreement between bone age stages using the CASAS and manual methods.

It has been claimed that the use of a continuous scale in the CASAS system leads to a smoother progression in bone age over time, with the standard deviations of the CASAS results being 50% of those for the manual TW method [Tann94b]. This is one of the reasons why the system has been recommended for the longitudinal assessments in children [Teun96], and the suggestion was even made that if patients are being further assessed, then their previous radiographs should be reassessed using the CASAS system [Tann94a]. In a series of 6-monthly serial radiographs of children, the stage of some bones appeared to reverse in approximately 4% of bones when using the manual method, but there were no reversals of one stage or more when using the CASAS system [Tann94a]. Another group found that in a limited sample of five patients with Turner's syndrome, three patients showed a reversal in bone age when assessed using the CASAS system, but there were no such reversals when they used the manual TW2 and GP methods [Fris96].

General opinions on the CASAS system seemed to vary, although the general conclusion was that it was probably adequate for bone age assessment in children with normal bone morphology. Some of the reported drawbacks with the CASAS system were that it worked best with high resolution radiographs typical of those pre-1990, and that non-standard hand positioning and unusually shaped bones caused poor matching to bone stages [Tann01, p24] [Fris96] [Teun96].

Overall, research on the user-assisted CASAS system has shown that it is possible to increase repeatability, and that it is possible to use relatively simple image processing techniques to perform the TW method of bone age assessment with reasonable accuracy. However, the system appears to fail when the bone age is assessed in children with some pathological conditions, most likely because of bone deformations in these conditions. The system can also require a large number of manual interventions, which reduces the objectivity of the assessment.

2.1.2.2 Analysis of the middle phalanx of the third finger using an active shape model: Niemeijer *et al.*

A limited implementation of the TW2 system was performed by Niemeijer *et al.* [Niem02] whereby they classified the stage of the middle phalanx of the third finger using an active shape model. An active shape model is an model of the mean object shape with and an added eigenvector description of the most significant variation modes for the shape– described using a covariance matrix. It is essentially an iterative, deformable model with shape constraints imposed using a statistical measure derived from a training set of contours. In this system, the user would draw a box around the third phalanx and the computer would automatically segment the enclosed bone using the active shape model. Classification of the bone stage was performed by finding the highest correlation between the pixel values for a physeal region of interest and the pixel values from a set of mean images representing the limited range of TW2 stages. Compared with a trained observer, their system had an accuracy of 73%, and this compared favourably with an accuracy of 80% for a second human observer [Niem02, p62]. However, this staging was only useful over the TW2 stages of E to I, corresponding to an age range of approximately 9 to years 17 . Furthermore, by using only one bone the precision of the estimates may be poor because, for example, the step between two stages in the TW method can be as high as four years in males [Tann01, p2].

2.1.2.3 Neural network-based system using linear distance measures: Gross *et al.*

The alternative to implementing a morphology-based system like the TW method is to have the user directly measure features from the radiographs and to use a decision system to calculate the bone age. The approach taken by Gross *et al.* was to have a user position cursors on digitised hand-wrist radiographs to measure distances. The ratios of these distances were feed into a neural network that had been trained to calculate the bone age [Gros95]. They started with ten measurements, but a linear regression analysis allowed them to choose seven measurements whose ratios against chronological age gave the best correlation coefficients for a group of male children. From these correlation coefficients they chose three ratios with the highest correlation coefficients. The drawback of this technique was that it ignored morphological information that is used in the GP, TW, and Fels methods. This may be one of the reasons why they found that all correlation coefficients for the ratios were less than 0.67. They found, however, that there was no significant difference between the bone age derived using their neural network and that assessed by a single radiologist using the GP method. They also found no significant difference between the mean neural network bone age and bone ages from 14 images from the actual GP atlas. These results show the ability of their neural network to find relationships between variables that each have a low correlation with bone age.

2.1.3 Fully automated assessment

Methods for fully, or near–fully, automated assessment of bone age are the topic of this dissertation. Although a number of researchers have worked on methods to fully automate assessment, few have published details of complete systems. This section provides a summary of the research on complete systems.

2.1.3.1 Phalanges, epiphyses, and carpal bones: Pietka, Gertych, Huang *et al.*

The research by Pietka, Gertych, and Huange *et al.* was a significant contribution to both methods of bone age assessment and the operation of a fully automated system. They developed a clinical system that integrated with a Picture Archiving and Communication System (PACS), and included a digital atlas of bone age radiographs that could be updated in a controlled manner [Piet05]. They successively refined their system since first publishing results in the early 1990s.

The general approach taken by this group was to first perform a rough estimate of bone age using phalangeal length measurements and phalangeal length tables developed by Garn - as discussed in [Piet91]. Based on this bone age estimate, their system would decide which features to extract from the various regions of the hand-wrist. For example, carpal bone analysis would only be performed if the bone age was less than nine years of age in boys, and less than eight years of age in girls [Piet93]. The approach of using the phalangeal length to make decisions on further analysis presents the question of what happens to the assessment if this initial estimate is inaccurate, especially at bone ages around the switchover to different methods of region analysis. This concern arises because it was initially found that accurate measurements of phalangeal length could be performed, but these were not a good indicator of skeletal maturity [Piet91]. The phalangeal length is a surrogate for height and is not a reliable maturity indicator.

The image processing methods they used to analyse and extract skeletal maturity features were relatively simple, yet surprisingly robust for the age ranges they analysed. Their objective in analysing maturity features of the bones was to find features that had a high discrimination power with bone age, but not necessarily those that were in clinical use [Piet95]. Their expectation was that features should be reliably extracted and should not require "too sophisticated and time consuming algorithms" [Piet95]. Although in many respects it ignores the wealth of knowledge that has been introduced through the selection of skeletal maturity indicators used in the GP, TW, and Fels methods, this is a pragmatic and robust approach.

Throughout most of their research they used a fuzzy reasoning system [Piet05, Piet03b, Piet97,Piet95]. The justification for the use of fuzzy classifiers was the ability of this method

to cope with inexactness, missing data, and subjectivity in the bone age assessment. This was especially important for the carpal region because the bones not only have changing features with age, but in younger children some of the bones are not even present. A potential limitation of this fuzzy classification system was that it was developed using relationships with age rather than a measure of skeletal maturity, making it dependent on the reference population used to develop the fuzzy classifier.

It is a little unfortunate that most of the results published by this research group focussed on the accuracy of their methods of region extraction and segmentation, with few assessments of the accuracy of bone age results. They did compare their method of bone age assessment with chronological age, showing that the average difference between the two was about a year [Piet05, Piet03b]. They also compared computerised phalangeal-derived and carpus-derived bone ages with those performed by a radiologist using the GP method. There were nearly three times as many carpus-derived cases than phalangeal-derived cases that differed from the GP measurements by more than one year [Piet95]. However, the overall difference between the computerised assessment and manual GP measurement was less than six months in 73% of phalangeal-derived cases and 63% in carpus-derived cases, demonstrating the reasonable performance using both of these methods.

With the integration of their bone age assessment system into a clinical PACS environment [Piet03b, Piet05], the research group were careful to highlight that such assessment is fundamentally only computer-assisted because a clinician is always responsible for the final diagnosis [Piet95]. To assist the clinician, they automatically presented a comparable bone age image from a standardised digital atlas, chosen using the bone age calculated by their system [Piet03b]. To do this, their system had web-based image distribution and a digital atlas based upon a structured query language engine that controlled data storage and retrieval [Piet03a]. They also used a special graphical user interface to view image analysis during development of the system and to analyse the cause of image processing failures [Piet04].

Personal correspondence and publications indicate that the research of this group is continuing in subgroups working on different issues to improve and extend system performance. This works includes combining Gibbs random field and active contour model segmentation methods to improve the segmentation of phalangeal epiphyses [Gert07a]; improving the clinical integration of the system and quality assurance [Gert07b]; expanding analysis of the carpal bones [Zhan07]; and expanding the age range of the system by including analysis of the radial epiphysis in older children [Tsao08].

2.1.3.2 The third digit - three epiphyses: Sato *et al.*

Using a simpler approach of analysing only the bones of the third digit, Sato *et al.* claimed that they could automatically assess the skeletal maturity of Japanese children [Sato99]. Their system was called the computer-aided skeletal maturity assessment system (CASMAS). It analysed the proximal, middle, and distal epiphyses of the third digit using ratios of the widths of the epiphysis to metaphysis, as well as the ratio of the metaphysis-epiphysis overlap to the width of the metaphysis. Where fusion of the epiphyses was nearly complete, they included the option of analysing the radial epiphysis.

For the three epiphyses of the third digit, and across the age range 2–15 years, they found that the percentage of successful automatic evaluations varied enormously. The results tended to be worse for the very young and older children, which can be explained by the under-development of the epiphyses in the young, and problems with bone overlap in older children. In those children for whom automatic evaluation of their radiographs was possible, there was a good correlation between the ratio of the epiphyseal-metaphyseal width and their chronological age. This finding of a good correlation between bone age and maturation of a limited set of bones is similar to another study that found high correlations between overall bone age and the individual bone age of the capitate, metacarpal III, proximal phalanx III, and middle phalanx III [Clar62]. Although such limited sets of bones may provide an indication of bone age, there is the risk that variability in bone maturity within this set will produce an unreliable bone age estimate [Roch88, p29]. When many bones are included in the assessment, the influence of variability in one or two bones will be less than if only a limited set of bones is used, depending to some degree on the method of estimating the bone age. Additionally, a concern that arises is whether such correlations between limited sets of bones and the overall bone age are valid in children with pathological conditions where bone age variation of individual bones is greater.

2.1.3.3 Phalanges, epiphyses, and carpals: National Tsing-Hwa University

The bone age research group at the National Tsing-Hwa University (Taiwan) initially focused on analysis of the third digit [Hsie07b] [Chan03], similar to that by Sato *et al.* (above). Their system included a method for extracting the left hand from a radiograph with both hands on the same film [Hsie07b]. Their preprocessing methods included thresholding techniques and heuristics searches that resulted in rotation of the hand image so that the third digit was vertical [Hsie07b]. A phalangeal region of interest (PROI) was found by scanning elliptic arcs over a fixed angle about an initial, approximate direction of bones of the digit. Segmentation of the phalangeal bones themselves involved a combination of Gabor filters to smooth the image, and edge finding and refinement using a Canny edge detector and a local variance method. The resulting binary image of the segmented phalanx

was used to mask the original image and produce a segmented PROI image containing full grey-scale information. Their segmentation method appeared to be very successful, with failure of the PROI extraction in only 2 from 360 female radiographs, and 5 from 360 male radiographs [Hsie07b].

Two sets of features were extracted from the segmented PROI images. The first was a set of geometric measures of the length, width, and area of the distal, middle, and proximal phalanx. The three regions were found by searching for minima in a profile of summed pixel values across the PROI. The areas and lengths were normalised using the length of the distal phalanx. The choice of the distal phalanx for normalisation was a little surprising because this bone is known to have poor contrast in some radiographs, so it can be difficult to accurately locate the end of the bone for a measurement.

The second set of features was based on the shape of the epiphysis of the distal phalanx. Horizontal projection profiles of both the binary image and grey scale image were derived from an analysis window that enclosed the metaphysis and epiphysis of the proximal phalanx, and the diaphysis of the middle phalanx. They used the discrete cosine transform to derive a frequency space representation of these profiles. The dimensionality of the features was reduced by choosing the first two significant components from a principal component analysis of the normalised lengths, areas, and nine frequency components from the discrete cosine transform. These components were feed into a neural network.

Three neural networks were investigated using the leave-one-out statistical method of training and testing. These were a back-propagation neural network, a radial basis function neural network, and a support vector machine neural network. Their support vector machine performed the best, but even that network only gave an accuracy of 85% when a bone age error of 1.5 years was allowed.

To improve the estimate, they decided to incorporate carpal bone information into the bone age assessment for children less than eight years of bone age (as assessed by the carpal analysis itself). The carpal bone age was calculated using a fuzzy membership function for the ratio of carpal bone area to total carpus area[2]. They did not provide details of how they calculated the carpal bone area. The carpal bone age was applied as a mask to neural network output values. For any given bone age less than or equal to seven years, a final bone age estimate was made using the neural network output corresponding to the carpal bone age estimate, plus the two outputs above and the two outputs below the estimate. This meant that five outputs were weighed about the output corresponding to the carpal bone age estimate. If the carpal bone age was more than seven years, the two largest neural network outputs were used. Hence the fuzzy estimate of carpal bone age

[2]In a recent publication this group further validated their use of the carpal bone area ratio applied to four different classifiers [Hsie07a]

was used to constraint the neural network result for the phalangeal processing. This fuzzy constraint, combined with the use of a support vector machine, resulted in a small increase in correct classification rate that demonstrated the usefulness of including the carpal bone information in the bone age assessment. In a recent publication this group further explored the use of a phalanx-based neural network system for classification of bones, with the bone age estimate confined by a fuzzy-based carpal bone analysis [Hsie08a].

This group is now extending their system to include screening for Turner's Syndrome by measuring both the bone age and a measure of the distal-middle phalanx ratio [Hsie08b], a ratio that often changes in this syndrome.

2.1.3.4 Phalangeal analysis using an active shape model: Mahmoodi *et al.*

The phalanges were also processed in an automated system developed by Mahmoodi *et al.* [Mahm00]. Their system first used a hierarchical search method to localise the bones, then a bone contour was found using an active shape model. From the contour of the active shape model they derived three bone shape features whose correlation coefficients with chronological age were between 0.72 and 0.89. The three features were the coefficient of a principal component from the active shape model; a shape moment of the proximal end of the phalanx; a ratio of the width of the epiphysis to the width of the metaphysis. They also looked at texture analysis of the epiphysis-metaphysis region and although they found a reasonable correlation between the results from this analysis and chronological age [Mahm98, p129], they appeared to dismiss this finding in a subsequent publication of their research [Mahm00].

Using the three features, they estimated the bone age by calculating an age that minimised a risk function based on the Bayes risk principle from decision theory [Mahm00] [Mahm98, p132]. The accuracy of this method was tested using a leave-one-out technique whereby one sample (patient radiograph) was removed, the system was trained using the remaining samples, and the removed sample was tested using new parameters based on the training. They claimed bone age accuracies of (82$\pm$3)% for males, and (84$\pm$3)% for females [Mahm00]. They also proposed that these accuracies could be further improved by using a larger training set that would allow them to model the variance in their sample population, rather than assume that the variance was the same across all age ranges (which is what they had done in this research).

2.1.3.5 Neural network classifiers for geometric features of the RUS and carpal bones: Liu *et al.*

A large patient database was used by Liu *et al.* to train and test an artificial neural network system that used geometric features of length ratios and area ratios of the RUS and carpal

bones to compute a bone age [Liu08]. Particle swarm optimisation was used with bone templates to segment the bones. The system used two classifiers: one for the RUS bones and one for the carpal bones (for children less than nine years old). Both classifiers produced very small mean bone age differences between the system and human observers, and a small standard deviation of the differences. An important finding from their work was a reduced variability in carpal bone-based bone age estimates made with their automated system. This potential improvement by automation is discussed in more depth in Section 2.3.

2.1.3.6 Neural network Analysis of the radius and ulna: Tristàn and Arribas

Tristàn and Arribas used neural networks to produce a bone age estimate using an implementation of the TW method for the radius and ulna [Tris08]. Their method included an adaptive clustering algorithm to segment the radius and ulna after manual assistance with some land marks. A unique aspect of their research was the use of neural networks in the decision stage that generated *a posteriori* probabilities that could be used to estimate the diagnosing error. Although the mean difference between their system and two human observers was small, the range of the differences were large, and their method was limited to only four TW3 stages for the radius, and five for the ulna. However, they believe their system could benefit from improvements in bone segmentation and further training of their neural networks. Their results suggest that neural networks are worthy of further investigation.

2.1.3.7 Neural network analysis of the epiphyses and carpal bones: Rucci *et al.* and Bocchi *et al.*

The usual approach to the task of analysing the bones of the hand and wrist is to generate an outline of the bone boundary and to extract features from either this outline or its interior. Finding the bone boundary can be difficult in the carpal bones because of low edge contrast, overlapping soft tissue, and noise in the radiographic image. Rucci *et al.* claimed that they could avoid this problem by training neural networks to look for image features [Rucc95]. They did this using a neural network architecture consisting of an attention focuser and a bone classifier. The attention focuser used what they called a 'retina', effectively a 'pixel-processing' neural network that fed into hidden neural network layers to produce an output of X and Y values corresponding to the centroid of each bone in the radiograph. The centroids were used to position square regions of interest over each bone. A classification neural network, using a similar retina structure, then classified the bone into one of the TW2 stages.

The network was trained using 56 relatively low resolution radiographs, and tested us-

ing another 16 radiographs. The gold-standard for the training was an assessment by an 'expert physician' using the TW2 method. They reported a bone stage classification accuracy of 65% for the correct stage, and 97% for a difference of less than or equal to one stage. The average difference between the bone assessed by the expert and their system was 0.62 years, with a standard deviation of 0.58 years. Once again, these results indicate the usefulness of neural networks in the classification task for the TW2 method.

Although not presented as a fully automated system, the research by Bocchi's group [Bocc03] [Bocc04] if combined could result in a system that would be almost fully automated. In their early work, a user manually selected regions from the radiograph, but in subsequent research they investigated a neural network method for locating and labelling the bones. They used similar 'pixel-processing' methods to those investigated by Rucci *et al.* [Rucc95], except their system included analysis of the epiphyses. For the manual method of bone localisation and labelling, they found an average difference of only 0.05 years, and a standard deviation of 0.7 years between the bone age measured using their system and that of an expert physician [Bocc03]. The maximum error was 1.4 years.

The results from both these groups appear very good. However, it was unclear over what age ranges they had tested their systems, and whether their systems would still work if a bone was missing, for example, before ossification in young children. Furthermore, it was unclear whether their systems would cope with bones that had started to overlap.

This neural network research demonstrates an interesting approach to storing processing information within a neural system. The conceptual difficulty with the approach used so far is that the neural systems start out dumb. In doing so, they effectively ignore the significant amount of research that has gone into developing skeletal maturity indicators for bone age assessment.

2.1.3.8 The Royal Orthopaedic Hospital Skeletal Ageing System: Hill and Pynsent

The Royal Orthopaedic Hospital Skeletal Ageing System (ROHSAS) undertook an automated implementation of the 13-bone and 20-bone TW2 assessment without any human intervention. Their system was able to complete an assessment in approximately four minutes using 1994 personal computer technology [Hill94b]. It used an iterative method to find the hand outline and identify the phalanges, carpus, radius-ulna area, and could identify whether the left or right hand was presented based on assessment of the widths of the radius and ulna. Segmentation of the bones was based on the fuzzy set and entropy method [Pal83]. The bones were classified using a shape recognition method based on normal, fuzzy, and fractional fuzzy grammars that were analysed using octal chain codes that described the edges of the bones [Kwab85]. As with the CASAS system, the user could

intervene and override the classification of individual bones, if required.

The ROHSAS system was evaluated by Cox using 98 radiographs from the International Children Centre London Longitudinal Study [Cox94]. In a comparison of bone age assessments using these radiographs, Cox found that the mean interobserver difference was 0.35 years, and 0.5 years between one observer and the ROHSAS. There was no difference in bone stage between the computer and a single human observer in 74% of the bones assessed. The ROHSAS assessed bones as one stage higher in 14% of the bones, and one stage lover in 11% of bones. However, the rejection rate of the system was 25%, most of which the researchers attributed to their system not coping with small distances between the edge of the hand the edge of the radiographic film. The conclusion by Cox was that the grammar-based system produced acceptable results, but the concern was that it required a much larger set of normal radiographs for validation, as well as evaluation against a set of radiographs outside of the normal range [Cox94].

2.1.3.9 The RUS bones analysed using BoneXpert: Thodberg *et al.*

The BoneXpert system is a recent and extensive addition to the research on automated bone age assessment [Thod09]. This system uses shape-driven active appearance models to segment the radius, ulna, and short bones (the TW RUS bones), ignoring the carpus. Active appearance models are similar to active shape models, but include a model of the intensity variation of a region - the shape and intensity variations constitute the 'appearance'. The parameters of the active appearance model are derived from a principal component analysis of a set of rotated, scaled, and registered bone contours. More than 3000 bone contours were used to derive the principal components for the shape variation, intensity variation, and texture (using Gabor filters)[3]. A set of coefficients are produced by fitting the active appearance model to a bone, and a selection of 30 of these coefficients were used as image features for a linear regression model of the age of a bone. The ages of the individual bones were then combined to give an overall bone age estimate. Additionally, the system was calibrated against the Greulich and Pyle atlas to give GP bone ages, and TW2 bone ages were calculated using a slightly unusual method that produced TW stages from individual ages of bones, then recombined them to give a TW2 bone age.

The performance of the system is still being actively evaluated, but initial results indicate that the performance is good, with an accuracy of 0.42 years (standard deviation) for the GP method, and 0.80 years for a cross-validated study using the TW2 method [Thod09]. The system has the ability to reject radiographs of poor quality. The overall rejection rate was very low (1%), although the rejection of individual bones was sometimes high (as high as 18% for the radius and ulna). An especially useful output of the research was a

[3]The shape analysis was similar, in principle, to that developed by Mahmoodi *et al.* [Mahm00]

measure of bone age precision based on an interpolation of the bone age from sequential radiographs, and assuming linear growth. The precision of the system was 0.17 years (standard deviation), and this was very good compared with an average operator precision of 0.50 years, demonstrating the improvement in reproducibility that can be achieved using an automated system.

The BoneXpert system has been released as a commercial product with pay-per-use licensing (http://www.boneXpert.com, Jan 2009).

2.1.4 Summary

Three approaches to the computerised assessment of bone age have been discussed: computer-assisted manual methods, user-assisted computerised methods, and fully automated methods. This review has demonstrated, through examples, that it is possible to achieve reasonable bone age accuracy with each of these approaches.

Some of the fully automated methods are limited in that they only use of a small set of bones in the analysis. If only a limited set of bones is chosen, such as the third digit method, then there is more risk that variability in bone maturity across the hand could result in an unreliable bone age estimate. The range of bone-specific skeletal ages within the hand and wrist can be large, even within normal children [Roch88, p29]. Furthermore, some pathologies can exaggerate bone age differences between the bones of the hand and wrist [Roch88, p29]. This presents a challenge for fully automated assessment: include as many bones as possible to span the period of skeletal development and to protect against variability in bone maturity, yet use this variability to indicate the confidence in the bone age estimate.

Most of the effort given to automated bone age assessment has been focussed on the radius, ulna, and short-bones – the so-called RUS bones. This review has shown that research on the automated assessment of carpal bones has been limited to younger children, less than nine years old. The role of the carpal bones in older children has not been investigated. This is a research opportunity that is discussed further in Section 2.3.

2.2 Model of a system for the computerised assessment of bone age

The aim of this brief section is to describe one possible model of a fully automatic system for bone age assessment, or at least a system with minimal expert-user intervention. Included is discussion on approaches that other researchers have taken to address requirements of this model.

2.2.1 Overview of the model

The task of automating the assessment of bone age from hand-wrist radiographs is mostly one of computer vision. A computer is presented with an input image that is processed to generate a description of the objects within the image (the bones) which is then used to derive a measure of the bone age of the child. One model for such a system is shown in Figure 2.1. The model depicts individual stages of the assessment process, but in a practical implementation the stages are often interdependent. For example, some methods require the extraction of a limited set of features before image segmentation can be performed.

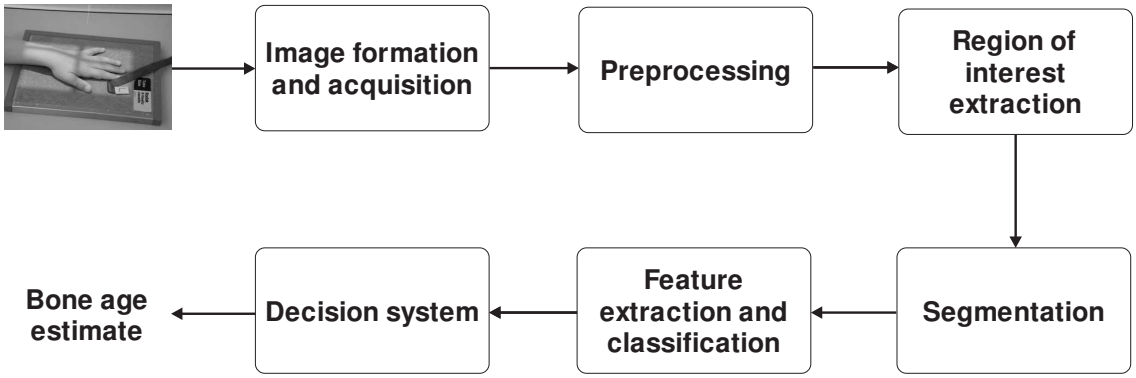


Figure 2.1 Model of the proposed bone age assessment system.

This is only one possible model. An example of another model is the use of 'pixel-processing' neural networks that attempt to circumvent the need for segmentation. Such techniques have been used by Rucci *et al.* and Bocchi *et al.* for bone age assessment (see Section 2.1.3).

The various stages of this model will now be briefly described.

2.2.1.1 Image formation and acquisition

The hand-wrist radiograph is formed from x-rays leaving an x-ray source, passing through the hand, and some of the transmitted x-rays causing ionisation in an x-ray sensitive receptor. The x-rays usually pass through the hand in a straight line. Of those that do not, some undergo scattering, and the rest are completely absorbed by bone or soft tissue. The amount of x-ray scattering and absorption depend on the energy of the x-rays and the thickness, density, and atomic number of the material through which they pass. It is the differences in these four parameters that cause differences in the relative x-ray flux at the x-ray receptor. For example, if the x-ray energy is high there will only be a small difference between the absorption of the x-rays in bone and soft tissue. This results in a loss of contrast between the bone and soft tissue on the radiograph.

The type of x-ray image receptor can also influence the contrast. For example, a receptor consisting of a combination of an x-ray sensitive, light-emitting screen with a light-sensitive film has a non-linear dependence on the flux of x-rays reaching the screen. The non-linearity in this screen-film combination is caused by the optical response of the film. If the x-ray flux is small (underexposure) or large (overexposure), there is loss in image contrast. The screen-film combination will only give optimal contrast performance over a limited range of x-ray flux. If the x-ray exposure is poorly controlled, the image contrast will be degraded.

To expand the optimal-contrast range, traditional screen-film combinations can be replaced by computed radiography or direct radiography. Computed radiography uses photostimulable phosphor plates that store the ionisation caused by the x-rays as a latent image, that is then readout from the plate using a laser beam to stimulate the phosphor. Alternatively, a direct radiography panel can be used to immediately convert the x-ray ionisation to an electrical signal for storage as digital image [Doi06]. Unlike screen-film combinations, both of these systems can control the relationship between the incident x-ray flux and the resulting output pixel values. This allows changes to the image contrast and dynamic range, and improves tolerance to over and under-exposure of the x-ray receptor. If required, spatial filters can also be applied to enhance edges.

With computed radiography and direct radiography systems there are no further steps in the image acquisition stage for computerised bone age assessment - the image is inherently in a digital form that is stored ready for transfer to a processing system. With screen-film systems, the film has to be digitised. Dedicated x-ray film digitisers are available for this task, but digitising video cameras have also been mounted over x-ray light boxes to capture the image data.

The digital form of the radiographic image is usually a rectangular array of pixels with depths of 8, 12 or 16 bits. The overall contrast in the image is determined by the energy of the x-rays, scattered and absorbed radiation in the tissues of the hand, energy-dependent absorption of x-rays at the image receptor, and the relationship between the x-ray flux reaching the receptor and the output pixel values. The overall image resolution is determined by the size of the x-ray source in the x-ray tube, the distance from the x-ray tube to the image receptor, the distance between the hand and the image receptor, and the design of the image receptor itself. Often only the image receptor resolution is considered because the other factors are controlled by a standard radiographic exposure (Section 1.2.1).

The noise in the digital image is determined by a combination of structural noise sources from the image receptor, variations in the x-ray fluxes due to stochastic processes of x-ray generation and interaction (quantum mottle), and quantization noise from the digitisation

process. Although not strictly a source of noise, the radiographic projection of the internal composition of the bones - the bony trabeculae - can produce image textures that are difficult to analyse.

2.2.1.2 Preprocessing

The preprocessing requirements usually depend on the subsequent methods used in the bone age assessment system. For example, if a segmentation algorithm is intolerant of noise, then significant spatial filtering may be required as part of the preprocessing. Typically, the preprocessing prepares the image for subsequent analysis. This may include removing or excluding image labels such as the 'L' typically used to indicate the left hand, as well as eliminating white borders caused by collimation of the x-ray beam, and reorientation of the image.

The preprocessing system should also be responsible for the initial screening of the hand-wrist radiograph. This should include rejecting poor quality radiographs that would lead to erroneous results. Such radiographs would then be flagged for either user intervention or full, manual reporting by a radiologist or specialist physician.

2.2.1.3 Region of interest extraction

The purpose of the region of interest (ROI) extraction stage is to break down the bone age assessment into different processes. The GP, TW, and Fels methods separately consider the phalangeal and metacarpal epiphyses, the carpal bones, and the radius and ulna epiphyses. The model adopted here is that all three regions should also be considered separately in an automated system.

The output of the ROI extraction stage is a numerical representation of regions surrounding the areas of interest. These regions usually completely enclose the bones of interest, but may focus on areas such as the metaphyseal-epiphyseal region.

2.2.1.4 Image segmentation

Image segmentation can be complex because there is no standard definition or theory for it, nor single standard method of performing it [Prat01, p551]. The segmentation process of this model refers to the separation on bone structures from other structures of the hand and wrist by way of some attribute of the bone. The attribute is commonly based upon differences in intensity in the digital radiographic image, but it could also be based on another attribute such as a region of common bone texture. The segmentation is usually described by a contour defining the edge of a bone, or region of bone. This contour may be open or closed, depending on the bone and the features that are extracted from it.

2.2.1.5 Feature extraction, classification, and parameterisation

The term image feature is sometimes used to refer to low-level characteristics or attributes of an image [Prat01, p509] and such features are used for the image segmentation stage. However, in this model feature extraction is the process of extracting higher-level image features from a segmented radiograph such as epiphyseal outlines, and these features have some relationship with skeletal maturity indicators. These features may be used to classify a bone into some stage of development, like that used with the TW method, or alternatively, the features may be described by a set of parameters that have a direct relationship with skeletal maturity.

2.2.1.6 Decision system

The decision system estimates the bone age either from a set of bone classifications, or from a set of parameters describing features of the various bones. The process of estimating a bone age is dependent upon a reference population and is implemented as a calibrated mapping from the bone classifications or parameters to a bone age estimate based on the reference population. The alternative approach is to estimate a skeletal maturity score that is independent of the reference population, as in the TW method, and to subsequently convert this to bone age using calibrated conversion factors [Tann01, p9]. Ideally, the decision system should also give an indication of the uncertainty in the bone age estimate.

2.3 Improving the computerised assessment of bone age – the inclusion of carpal data

Although there has been a significant amount of research on the computerised assessment of bone age, especially with respect to the phalanges, there remain gaps in knowledge and opportunities for improving the assessment. An example is the lack of research on the role of the carpal bones and the contribution they can make to computerised assessment. It remains unclear whether incorporating carpal bone maturity indicators into a computerised system can improve the accuracy of bone age assessment in older children. The carpal bones have a wide range of morphology and in older children they begin to overlap, which adds extra complexity to their automated analysis. Furthermore, there is concern that including the carpal bones in bone age assessment has a negative influence upon the accuracy and precision of the assessment because of interobserver and intraobserver variability, and inherent differences in development between the carpal bones and the phalanges [John65].

The objective of this research is to develop methods for the computerised analysis of carpal

bone skeletal maturity in children over nine years of age[4]. The aim is to incorporate this analysis into a system for the automated assessment of bone age. The motivation for the research was the hypothesis that if interobserver and intraobserver variability are the major contributors to uncertainty in bone age estimation, then automating the assessment should decrease this uncertainty and provide a more objective measure of the contribution of carpal bones to bone age assessment. In addition, improved precision and accuracy of carpal bone analysis could improve the precision and accuracy of the overall bone age estimate, the importance of which was discussed in Section 1.2.3.5.

Therefore, this final section of Chapter 2 introduces three areas that require investigation: segmentation of the carpal bones, extracting features from the carpal bones, and incorporating carpal bone features into a decision system for bone age estimation. The discussion of the basis for this research begins by reviewing what has been established about the role of the carpal bones in bone age assessment.

2.3.1 The role of the carpal bones in computerised assessment

It is proposed that the carpal bones of older children contain indicators of skeletal maturity that have not been utilised in systems for the computerised assessment of bone age. The research question is whether incorporating these indicators into a computerised system would improve the assessment of bone age in children over nine years of age.

2.3.1.1 Should the carpal bones be included in bone age assessment?

In addressing this question, it is important to consider both the contribution the carpal bones make to accuracy, and the impact including them has upon the precision of bone age assessment. This, however, appears to be a difficult question to answer because a review of the scientific literature shows conflicting claims about the role of the carpal bones in bone age assessment. The reason for this difficulty is most likely a combination of differences in sample populations for research studies, and differences in the methods of assessment - with corresponding differences in accuracy and precision.

Comparison of assessments when carpals are included or excluded
The impact that including the carpal bones has upon accuracy can be determined by comparing bone age measurements with and without the carpal bones as part of the analysis. Unfortunately, this only provides a relative measure of the impact of the carpal bones because an absolute comparison would require prior validation of the accuracy of the bone

[4]The carpal bones begin to show a decreasing contribution to the bone age assessment in children beyond the age of nine years (Section 2.3.1.3). Additionally, the analysis is made more complex because the carpal bones begin to overlap. These are the two reasons why automated assessment in children over the age of nine years has received little attention

age method against the sample population used for the comparison.

Furthermore, this assumes the reference population matches the test population (as discussed in section Section 1.3.3) especially the concordance between carpals and the other bones. The accuracy of the GP method can be tested by simply excluding the carpal bones when comparing radiographs with the standards from the GP atlas. In a study using two observers, the difference between including and excluding the carpal bones was significant for all duplicate readings for boys and girls (a total of four readings), but for a second observer only one of four sets of readings were significantly different from each other [John65]. Given this inconsistency and that the maximum difference between including and excluding the carpal bones was only 3.9 months, it is difficult to conclude that their findings were significant. Another study used the GP atlas to separately measure bone ages for the carpals, wrist, and hand (phalanges and metacarpals) [Carp93], and found that the carpal bone age was 4.1 months less than the hand bone age in boys, and 8.1 months more than the hand bone age in girls. The influence of ethnicity on these differences was studied by Levine, who found that in White, Negro, Coloured, and Indian males the carpal bone age lagged the hand bone age an average of -6.25 months to -3.68 months, but for females in the same ethnic groups the difference ranged from -0.37 months to +4.12 months [Levi72]. Although Levine's results for boys are comparable with other studies [John65] [Carp93], there is poor consensus on ethnicity effects for females, an inconsistency that remains unexplained. However, it may indicate that different investigative groups have used slightly different criteria when comparing subjects and atlas images, since the GP method is subjective.

The TW2 and TW3 methods both have separate radius-ulna-short bone (RUS) and carpal bone ages that simplify relative measurements; hence, TW is probably the best method for comparison of results with and without the carpal bones. However, the very consistency of the TW method may be due to the discarding of subjectively variable, but useful, data, and this discarding could bias such an investigation. Even so, the mean difference between the TW-RUS and TW-Carpal scores was found by Tanner *et al.* to be zero for boys in the age range 2-13 years and zero for girls in the age range 2-11 years [Tann01, p12]. However, in another study the TW2 RUS bone age was approximately 1 year in advance of the carpal bone age [Cole88]. This latter study included boys and girls over 13 years of age, so caution is required when making a comparison with the findings by Tanner *et al.*. A third study of 4-12 year old children from Argentina showed TW2 Carpal bone ages lagged behind TW2 RUS bone ages by a year or more, although Tanner *et al.* suggested that this might have been due to the carpal bones not sharing the "secular trend universally seen in the RUS set of bones, rather than indicating a population difference"[5] [Tann01, p13].

[5]Secular trends are slow trends that are seen over long timeframes. In this case, the trend has been towards advancement of maturation and, hence, bone age. The suggestion by Tanner *et al.* was that this trend may have

Overall, this evidence for the impact of the carpal bones on the accuracy of bone age assessment remains inconclusive for both the TW and GP methods. It is clear, however, that the experts in bone age assessment cited above expected that including the bones would provide a benefit.

Comparison of intra- and inter-observer variability when carpals are included or excluded
Observer variability has received the most attention for its influence on the precision or reproducibility of bone age assessments, as mentioned in Section 1.3.1. One study using the GP method looked at between-observer differences, between-technique differences for bone age assessment with and without inclusion of the carpal bones, and intra-observer differences. They concluded that the carpal bones may provide some useful information, but that their inclusion did not assist, and probably hindered, bone age assessment [John65]. There are three concerns about the evidence used to support this conclusion: 1) the impact that including the carpal bones had upon systematic errors in the bone age assessment was statistically significant in seven sets of measurements associated with one observer, but only in one set of measurements associated with a second observer; 2) the between-observer differences were significant in two sets of measurements for girls, but one in favour of carpal bone exclusion, and the other in favour of carpal bone inclusion; and 3) assessment of variances between including and excluding the carpal bones showed no significant difference for both girls and boys. It is difficult to find a clear relationship between the evidence in this paper and its conclusions. Furthermore, another similar investigation concluded that excluding the carpal bones did not systematically affect either the interobserver or intraobserver variability [Roch70b]. It is unfortunate that despite these concerns, the earlier paper [John65] has been cited as a basis for excluding the carpal bones from computerised analysis in older children [Zhan07] [Piet97] [Piet93].

The lack of agreement on whether to include the carpal bones has been summarised by Roche *et al.* [Roch88, p15], who noted findings of variability in the rate and patterns of carpal maturation and noted that some researchers had concluded from this that the carpal bones should be excluded. However, they believed that this variability in the rate of maturation could be useful [Roch88, p16], and for this reason included a range of carpal maturity indicators in the Fels method, including typical variations in bone morphology at different stages of maturity. Furthermore, they looked at the contribution of the carpal bones to the assessment and found that for the Fels method the carpal bones contributed 40% of information for bone age assessment in the age range 5 - 9 years, and 20% for each of the ranges 3 - 5 years and 9 - 11 years [Roch88, p260]. However, their investigation of variability of grades showed that 52% of the intraobserver variability and 29% of the interobserver variability could be assigned to the carpal bones [Roch88, p269]. These findings highlight the problem with the carpal bones: the carpal bone contribution to bone age assessment is

only occurred for the RUS bones, and not in the carpal bones.

high in young children, but so is the intraobserver and interobserver variability associated with them. If a computerised system could reduce observer variability, then it is possible that the carpal bones could have a significant role in such a system.

Summary
There is no conclusive evidence to support exclusion of the carpal bones from bone age assessment. This contrasts with the evidence based on large reference populations that was used to support including the carpal bones in the GP, TW, and Fels methods [Greu59] [Tann01] [Roch88]. These methods demonstrate that the role of the carpal bones is well established in bone age assessment. For example, the TW2 20-bone method attributes 50% of the bone age assessment to the carpal bones [Tann83, p5]. In summary, the carpal bone assessment should remain part of the method for bone age assessment, and it should be a part of a computerised system for bone age assessment.

2.3.1.2 Using carpal bones for differential diagnosis

The argument for and against including the carpal bones has mostly focussed on the impact they have on bone age assessment in normal children. Some researchers have suggested that the difference in skeletal maturity between the carpal bones and the rest of the hand could be of use in the differential diagnosis of disease [Tann01, p12] [Kirk98]. It has been proposed that differences in development between the carpal bones and long bones could be due to different underlying hormone processes [Moli04]. Investigation of this idea, however, has been limited to specific diseases. In a study of children with a disorder that caused bone maturation acceleration, the TW2-RUS bone ages were at least one year in advance of TW2-Carpal bone ages, while the average difference was only 0.1 years for normal children [Vejv81]. Another study of children with renal disease also showed that TW2-RUS bone ages were an average of 10 months in advance of TW2-Carpal bone ages [Cund88]. It is possible, therefore, that bone age assessment could benefit from trying to separate skeletal development for the different regions of the hand, rather than trying to combine them into a single measure. In this respect, even if the carpal bone ages are shown to have a poor accuracy, they may be of use for differential diagnosis of disease.

2.3.1.3 Including the carpal bones in computerised assessment of older children

In children over the age of nine years, carpal bones frequently overlap, bone changes slow, and maturity indicators generally become more subtle and difficult to determine. This is the reason why the carpal bones are often not included in bone age assessment of children over nine.

Differences between the manual and computerised approaches to carpal bone assessment
The argument presented so far for including the carpal bones in computerised assessment

is: a) the GP, TW, and Fels methods all make extensive use of these bones; b) the evidence for excluding them is poor; and c) the carpal bones may assist differential diagnosis of disease. However, the idea that evidence from the manual methods of assessment can be applied to computerised methods may be incorrect and needs to be challenged.

Although human observers are very good at many of the visual and interpretive tasks required for bone age assessment, it is reasonable to expect that computerised methods will eventually be able to perform such tasks with at least equal ability. Because computerised methods have the potential to eliminate observer variability by elimination of random variations in the assessment, they may even prove to be better than human observers, especially in the precision of assessments.

There is the additional possibility that some quantitative measures of skeletal maturity that are difficult for human observers to perform could be of value when bone age assessment is computerised. For example, relative area measurements and the quantification of bone margin irregularity could lead to an improvement in the accuracy of bone age assessment.

Computerised assessment would also allow more advanced algorithms to be used for the decision system that derives a bone age from skeletal maturity indicators. An example of this is the use of computerised statistical models for processing skeletal maturity indicators in the Fels method [Roch88, p58].

Finally, implementation of computerised carpal bone assessment will allow further research to clarify the relationship between the isolated carpal bone age and skeletal maturity of both the hand and the rest of the body.

To include or not to include carpal assessment in older children - morphological considerations
The carpal bones in older children begin to take on their final form and do not change much. Once they reach adult maturity they make no further contribution to bone age assessment. At this point it becomes difficult to include them in such methods as the GP method, where bone-specific ages are averaged to give a bone age estimate [Roch88, p15]. This is because the skeletal development based on observation of these bones appears to have finished, and the lack of change in skeletal maturity prevents monitoring of growth. What is unclear is the level of maturity at which this occurs.

The GP, TW, and Fels methods all have carpal bone skeletal maturity indicators that progress beyond nine years of age. The Atlas of the GP method provides details of changes in the carpal bones until 11 years of age in girls and 12.5 years of age in boys [Greu59]. With the exception of the pisiform ossification, the Fels method requires that at least one maturity indicator should be analysed in all carpal bones until the age of 14.5 years in girls and 15.5 years in boys [Roch88]. The TW3-Carpal scores keep changing in girls until 11 - 15 years,

and until 13 - 17 years in boys (the ranges being the 97th to 3rd percentiles) [Tann01, p12]. Perhaps the most useful indication of the potential contribution of the carpal bones is the mean increment in the TW3-Carpal bone score per year. Results from Tanner *et al.* showed that the mean increment per year reached a maximum at approximately 8.5 years in both girls and boys [Tann01, p15].

These results are similar to the findings of Molinari *et al.*, reproduced in Figure 2.2 [Moli04]. Despite there being a rapid decline in the mean increment of the carpal bone score beyond 9 years of age, Figure 2.2 illustrates that even children at the 10th percentile (lower dashed line) show changes in scores to at least 11 years of age. These results do not suggest that maturity changes in the carpal bones are a reliable indicator of skeletal development, or that they would improve the accuracy of bone age assessment, but they do suggest that nine years of age is not the level at which such analysis should automatically stop.

To include or not to include carpal assessment in older children - clinical significance
In some respects, questions about whether the carpal bones should be included become more significant when considering their role in a computerised assessment system. Analysis of the carpal bones is complex and a decision has to be made about whether to invest the effort to include them. There has been a resurgence in interest in automated carpal bone analysis, but this has been limited to children less than nine years old [Lin04] [Zhan06]. The role of the carpal bones in the computerised assessment of older children is unknown.

Although this role is an interesting research question to address, and is stimulus for investigating the computerised assessment of these bones, equally important is the more immediate need for improved accuracy of bone age assessment in this age group. Puberty is an important stage of development of a child, yet there is limited consensus on when the GP method is least accurate, with some suggesting the early puberty period 10–14 years [Zeri91] and others the late puberty period 14–17 years [Kemp99] .

Finally, as introduced in Section 1.1.3, bone age monitoring is useful at all ages while growth hormone is being administered, but in older children the bone age assessment will also assist the decision to cease growth hormone therapy.

2.3.1.4 Summary

This section has presented an argument for why carpal bones should be included in bone age assessment. It is argued that even if the evidence for excluding the carpal bones is considered reasonable, this evidence may not be applicable to the computerised assessment of bone age. The potential benefits from computerised analysis are sufficient to warrant investigation of the role of the carpal bones in the bone age assessment of children over nine years of age. The rest of this chapter introduces the basis on which this could be achieved.

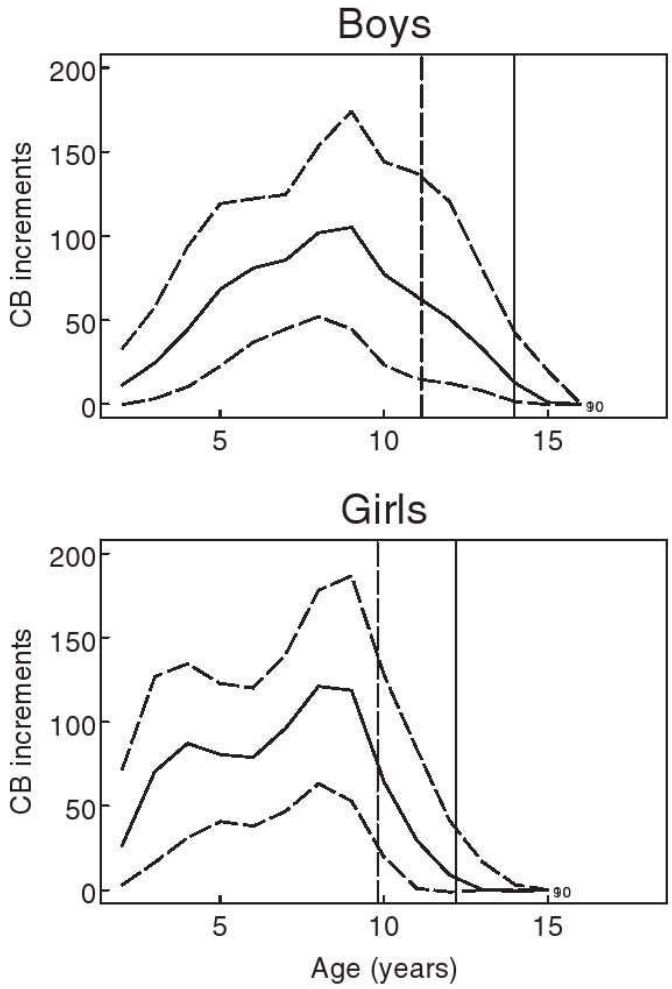


Figure 2.2 Mean increments in TW3-Carpals scores per year. The curves are the 10th, 50th (solid line), and 90th percentiles. The solid vertical line is the time of peak height velocity during the pubertal growth spurt. (reproduced with permission from [Moli04])

2.3.2 Segmentation of the carpal bones and the overlap problem

The problem

The methods for segmentation of the carpal bones on radiographs of older children are likely to be different from those for young children because around nine years of age the carpal bones begin to overlap, rendering the segmentation process more difficult. Unlike the problem of occlusion that has been extensively studied in computer vision, the overlap in radiographs is of structures that are semitransparent. This means that for two overlapping bone objects each include the other bone's edge within their boundary.

Segmentation procedure outputs

Early in the research for this thesis it was recognised that the complexity caused by the semitransparent scene, combined with possible poor radiographic definition of a bone edge, would mean that it would not always be possible to find a complete contour during carpal bone segmentation. Where components of a contour are too poorly defined to be of use, a mechanism is required to label these components, to allow them to be excluded from subsequent feature extraction and bone age assessment procedures, but to be included in measures of confidence in the segmentation.

2.3.3 Extracting features from the carpal bones

Overview

The GP, TW, and Fels methods all describe carpal bone skeletal maturity indicators that can be of use in older children (Section 2.3.1). It should be possible to automatically extract image features that relate to these indicators and hence have a relationship with bone age. Furthermore, rather than use feature extraction methods to classify a bone into a given stage of development, it should be possible to parameterise the features and provide a continuous scale for the maturity of an individual bone.

Maturity as a continuum

If the resolution used in the quantisation of the bone maturity stages is a significant component of the bone age error in methods like TW, then using a continuous scale should improve the precision of bone age estimation. The motivation for this is the observation that an error of as little as one stage in the TW analysis can change the bone age by more than 0.3 years [Alba95]. One group attempted to solve this by developing a continuous scale for computerised scoring of bones based on a distance measure of Fourier coefficients of the bone images compared with the coefficients of a set of template images [Tann94b]. But, as discussed in Section 2.1.2.1 their are serious shortcomings to this method. An alternative approach is to use a continuous scale to describe complex image features that have a closer match with the skeletal maturity indicators of the GP, TW, and Fels methods.

Feature selection

There are many features that could be derived from a two-dimensional image of a carpal bone, but only a few are likely to have a dependent and reproducible relationship with bone age. If a feature is related to or considered a useful skeletal maturity indicator, then there are two constraints upon the feature: 1) it should be independent of the physical build of the child, and 2) the endpoint of the feature should be common to nearly all children (see Section 1.1.2.2). However, it is possible that two image features do not meet these criteria, but there may be a useful relationship between them that does. An example of this is the epiphyseal and metaphyseal diameters of the phalanges. The diameters themselves

depend on the size of a child's hand (physical build), but the ratio of the diameters can be used to indicate the stage of maturity of the phalanges [Tann01, p72]. By investigating features that are independent of the physical build of children, it may be possible to reduce the variance of the feature across the population and improve its usefulness.

Much of the research on computerised assessment of the carpal bones has focussed on simple, geometric and gray-scale measures that are applied equally to bones in children less than nine years of age. Although such methods can be applied to the carpal bones of older children, in many respects they ignore the expertise that has been incorporated into the GP, TW, and Fels methods. It can be argued that simple features like ratios of areas are difficult for human observers to quantify, but easy to be incorporated into a computerised system. However, it can equally be argued that more complex features are also required to quantify skeletal maturity indicators in older children, an example being investigation of features to describe the development of articular facets of the capitate bone [Tann83, p73]. Neither the simple nor the complex features should be ignored. They both require investigation to see if incorporating them into a computerised system can improve the assessment of bone age.

2.3.4 A decision system that considers data confidence

For the decision system there are three areas in addition to current methods that should be considered for a clinical bone age system.

Missing data
As well as providing a measure of confidence in the carpal bone segmentation and subsequent feature extraction, it is important that the decision system also includes this confidence measure in the estimate of bone age. If the segmentation or feature extraction are so poor that a bone feature can not be included in the assessment, the system should generate a measure of uncertainty for the estimate.

Reporting measurement error
The decision system should reject as few patients as possible. However, there is a trade-off between rejections and accepting results that are inaccurate. An improvement in bone age decision systems would be the ability of a system to report when difficulty was experienced with the automated analysis, and either allow manual intervention by a user, or produce a bone age estimate with a warning about the limitations of the estimate.

Reporting individual BA variation
The ability to highlight discordant results would be an additional benefit. The highlighting of discordance between the maturity of the carpal bones and the rest of the hand and wrist[6]

[6]This also applies to highlighting discordance within groups, such as within the phalanges and within the

would not only provide an indication of confidence in the bone age estimate, it could be used in the differential diagnosis of growth disorders (Section 2.3.1).

2.3.5 Image database requirements

It is proposed that the ideal evidence that an image feature is a useful maturity indicator should come from analysis of sequential radiographs in a longitudinal study. A large collection of such radiographs should clearly demonstrate age-dependent changes in skeletal maturity indicators. This data would allow investigation of the predictive value of features derived from the carpal images, both individual features and groups of features. The major advantage of this approach is that for any single child it would be possible to determine when a given skeletal maturity indicator no longer shows changes with skeletal development. However, such a study would be difficult to perform because it would be necessary to follow children for many years.

The alternative is to perform a cross-sectional study of a range of normal children, or children assumed to be normal because they had radiographs that were not for developmental disorders. This is essentially a snapshot of the population at a given point in time, with a range of bone ages included in the study group. The major disadvantage of this method is that it is not possible to follow skeletal maturity indicators to the point at which they no longer change, and, hence, determine the predictive value of an image feature.

The original intention of this research was to use radiographs that were used for the development of the CASAS System (refer to Section 2.1.2.1). With the permission of Professor J M Tanner, Professor R D Bock kindly supplied 98 of the original images as 8-bit TIFF files scanned with a 50-micron pixel size. On close inspection by a radiologist, Dr Michael Hurrell, few of these images were considered of sufficient clinical quality to estimate bone age using the GP method. It was decided that this data set did not constitute a cross-sectional study and could not be used for determining the reliability of extracted features.

Permission was granted by the Upper South B Regional Ethics Committee (URB/06/05/032) to extract a limited collection of anonymised hand radiographs from the Christchurch Hospital Picture Archiving and Communications System (PACS) for a period of one year. This data was sufficient for the proof-of-concept research undertaken for this thesis.

2.4 Conclusions

This review has shown that the research to-date on computerised bone age assessment has largely ignored the role of the carpal bones in older children. When analysing the carpal

carpals.

bones of older children there are three areas that require further investigation: segmentation of overlapping bones, extraction of features from the bones that relate to skeletal maturity indicators, and optimising the use of these features to estimate bone age. The first of these, segmentation of overlapping carpal bones, is fundamental to the success of investigating the role of the carpal bones. However, equally important is the removal of noise from the radiographic images. The success of segmentation is closely linked with regularisation or smoothing of the image data through denoising. Such denoising is the topic of this thesis.

Chapter 3

Mathematical terminology and notation

This chapter introduces some of the mathematical terminology and notation used in this thesis.

3.1 Images and image derivatives

The hand-wrist radiographic images processed in this thesis are rectangular images with integer pixel intensities in the range 0–4095. These images are a discrete representation of the radiographic exposure incident upon the computed radiography plate. It is assumed that the discrete sampling size is small enough so that the pixel intensities can be approximated by continuous or piecewise continuous mathematical functions. The integer pixel intensities are treated as positive real numbers for the continuous mathematical formulation. Therefore, letting Ω be the closed spatial domain of definition of the image, the *image* I can be defined as[1]

$$I(x, y) : \Omega \to \mathbb{R}, \text{ where } (x, y) \in \Omega \subset \mathbb{R}^2.$$

The *image gradient* ∇I is the derivative of the image I with respect to the spatial coordinates and is a vector representing the direction and magnitude of maximum variations in the image

$$\nabla I = (I_x, I_y)^T$$

where I_x and I_y are the first partial derivatives of I with respect to x and y respectively.

The *directional derivative* of I in the direction $\mathbf{v} \in \mathbb{R}^2$ is given by the dot product of the image

[1] This notation is similar to that used by Tschumperlé [Tsch02].

gradient and $\mathbf{v}$

$$I_{\mathbf{v}} = \frac{\partial I}{\partial \mathbf{v}} = \nabla I \cdot \mathbf{v}.$$

In practice the image gradient is often computed following Gaussian smoothing of the image data. Because of the commutative property of the derivative and Gaussian smoothing, this operation can be combined as convolution with a Gaussian derivative function. This defines the gradient operation at a scale set by the spatial spread (σ) of the Gaussian smoothing.

3.2 Image processing using partial differential equations

In the last couple of decades partial differential equations (PDEs) and the calculus of variations have emerged as powerful tools for image processing. As the theoretical and numerical foundations of these methods develop, the drive is towards introducing new image processing tools and establishing the connection between these and existing image processing methods such as stochastic and wavelet methods [Chan05, p21].

Most of the PDE-based methods used for image processing can be divided into two main categories: energy-based methods and direct PDE methods. In image denoising the energy-based methods are an inverse problem solved through the minimisation of an integral equation that relates the image data with some approximation based on the desired denoised image. The use of the term 'energy' may appear a misnomer, but it reflects the strong connection of these methods with the physical sciences, and the Bayesian framework in which a likelihood is maximised by minmising its associated posterior 'energy' [Chan05, p18]. The integral or variational form can often be solved using calculus of variation techniques such as forming Euler–Lagrange equations for the functional that forms the integral. For example, if $I(x)$ is an original clean image and $I_0(x)$ the image corrupted by noise, then the Rudin, Osher, and Fatemi's total variation model is given by

$$E(I) = \frac{1}{2}\int_{\Omega} |I_0 - I|^2 dx + \lambda \int_{\Omega} \Psi(|\nabla I|)\, dx$$

where $E(I)$ is an energy functional associated with the image function I, λ is a positive bias weighting factor, and Ψ is a convex regularisation function that determines the smoothness of the final solution [Aube06, p70]. In this variational formulation $I(x)$ is can be found by solving the partial differential equations that arise from setting $E'(I) = 0$ (assuming E is convex). The solution is usually found through an iterative process, so that for iteration t the function I_t is a denoised image that converges to I, if certain conditions are met.

In the other case of the direct PDE methods, the PDEs are solved directly without intro-

ducing the idea of an energy. It is the direct PDE formulations that have been inspired by physical processes such as diffusion, advection, and transportation. The bonus of adopting such methods is that they have been present in the scientific community for some time and have a strong theoretical and numerical basis. For a solid mathematical foundation relevant to the use of PDEs in image processing, the reader is referred to some very good monographs [Aube06], [Chan05], and [Sapi01] (to name just a few).

3.3 Euler–Lagrange equations

Often when applying energy-based methods to image processing, the energy functional E is defined as an integral of some function f of the image I and image gradient ∇I. The general form of the energy is defined by

$$E(I) = \int_{\Omega} f(x, I(x), \nabla I(x))\, dx. \tag{3.1}$$

If the derivative f' satisfies some 'mild' growth conditions, and I and ∇I can both be differentiated (with both I and ∇I having finite integrals over Ω – that is, they are functions from a Sobolev space), then the Gâteaux derivative of $E(I)$ is given by

$$E'(I) = \frac{\partial f}{\partial I}(x, I, \nabla I) - \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial \xi_i}(x, I, \nabla I)\right), \tag{3.2}$$

where $\xi = \nabla I$ and n is the number of dimensions of I [Aube06, p37].

Setting the Gâteaux derivative $E'(I) = 0$ generates an Euler–Lagrange equation.

These Euler–Lagrange equations are PDEs that need to be solved using numerical methods such as finite differences, finite elements, or spectral decomposition. Finite difference methods are popular in image processing because images are usually formed from uniformly distributed pixels. Appendix A of [Aube06] provides an introduction to finite difference methods that are relevant to image processing.

3.4 Mathematical notation

$\mathbb{R}$	real numbers
$\mathbb{Z}$	integers, for example, the set of positive integers $\mathbb{Z}^+$
$\approx$	approximately equal to
∞	infinity
$\propto$	proportional to

$\sum$	summation
$\|.\|$	magnitude
$\forall$	for all
$\in$	element of
$\subset$	sub-set of
$\neq$	not equal to
$O(.)$	order
$*$	convolution operation
δ_i	small neighbourhood of i
$S \backslash i$	S excluding i
(i, j) or (x, y)	spatial location pairs, x and y
$I(i, j)$, $I_{i,j}$ or $I(\mathbf{x})$	Image intensity value (pixel intensity/value) at spatial location (i, j) or $\mathbf{x}$
$a_{i,j}^k$	discretised variable a at pixel location (i, j) after k iterations
Ω	rectangular image domain, $\Omega \subset \mathbb{R}^2$
Ω_i	region i of the image domain, $\Omega_i \subset \Omega$
$\partial\Omega_i$	boundary of the region Ω_i
∂_x	partial derivative with respect to x; and abbreviation for $\frac{\partial}{\partial x}$
∂_{xy}	abbreviation for $\frac{\partial^2}{\partial x \partial y}$
I_x	derivative of the image I with respect to the variable x
∇f	gradient of f
Δf	Laplacian of f: $\Delta f = \sum_{i=1}^{N} \frac{\partial^2 f}{\partial {x_i}^2}$
$\mathrm{div}(f)$	divergence of f: $\mathrm{div}(f) = \sum_{i=1}^{N} \frac{\partial f}{\partial x_i}$
$\mathbf{H}$ or $\nabla^2 I$	Hessian of I, in the classical sense [Aube06]: $(\nabla^2 I)_{i,j} = \frac{\partial^2 I}{\partial x_i \partial y_j}$
μ or $\bar{x}$	mean of a population or mean of a specified random variable
$\mathrm{E}\{.\}$ or $\langle . \rangle$	expected value operation (mathematical expectation)
σ	standard deviation (square-root of the variance)
$\mathrm{Std}\{.\}$	standard deviation operation
s	standard deviation (s^2 is an unbiased estimator of σ^2 the population variance)
$\mathrm{Var}\{.\}$	variance operation
$p(.)$	probability operator
$p(x\|y)$	probability of x given y
G_σ	Gaussian kernel with standard deviation or spatial spread $\sigma > 0$. For a 2D Gaussian kernel $G_\sigma(x) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right)$, where $x \in \mathbb{R}^2$.
$H(.)$ and $H_\epsilon(.)$	Heaviside function and regularised Heaviside function

$$H(x) = \begin{cases} 1, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0; \end{cases}$$

$\delta(.)$ and $\delta_\epsilon(.)$	Dirac delta function and regularised dirac delta function
$\|.\|$	norm of a matrix or vector (Euclidean norm, unless stated otherwise)
$\mathrm{trace}(.)$	trace of a matrix: $\mathrm{trace}(A) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii}$
$log(.)$	logarithm base 10 (common logarithm)
$ln(.)$	logarithm base e (natural logarithm)

Chapter 4

Pixel intensity and noise characteristics of the Kodak computed radiography system

4.1 Introduction

Unlike screen-film radiography, computed radiography has a wide dynamic range. This dynamic range means that radiographs are seldom overexposed or underexposed. However, to produce images that are suitable for diagnosis the CR systems employ post - processing methods that map the pixel intensities to display important radiographic detail. The algorithms used to adjust the computed radiograph data can significantly affect the pixel intensity and noise characteristics.

The system for computerised assessment of bone age introduced in Chapter 2 should be able to adapt to any image that is processed. The approach taken in this research, however, was to focus on processing of radiographs taken using the combination of Kodak Ektascan HR plates and Kodak Directview CR900 plate reader. The hand-wrist radiographs used in this research were from the Christchurch Hospital Picture Archiving and Communication System (PACS) and the majority of these were taken using the Kodak Ektascan HR/CR900 combination.

The purpose of this chapter was to determine whether pixel intensity and noise variations of the Kodak CR 900 system can be modeled and predicted, with the intention of producing working models that can be used in subsequent image processing tasks. This is in contrast with other theoretical models that more accurately reflect the performance of

the CR system, but are too complex to be used for image processing tasks. Furthermore, most of these more accurate models have only be validated for processing of the CR images in non-clinical radiographic modes (minimal processing). The approach taken in this chapter was to use a simple model of radiographic noise – one in which Poisson noise sources dominate – and to determine the applicability of this model for clinical hand-wrist radiographs. A simple model was used because it was anticipated that a signal-dependent denoising method would not require a precise noise model. It is important to investigate whether this model can be applied to the hand-wrist clinical radiographs because there are multiple stages in the Kodak CR image processing, such as Perceptual Tone-Scaling and Enhanced Visualization Processing [Gall10], each of which can affect the image noise. The focus is on the Kodak CR system, but the theoretical model and many of the investigative methods used in this chapter can be applied to other Computed Radiography and Digital Radiography technologies.

To streamline the presentation of arguments and findings, the experimental methodology is detailed in Appendix B.

Please note: With the exception of the following dose response relationship, all pixel intensities used in this thesis were for the 12-bit photometric interpretation of white on black. Under this interpretation the blackest pixels have pixel intensities of approximately 0 and the whitest pixels have pixels intensities of approximately 4095.

4.2 Pixel intensity and contrast characteristics

Three methods were used to investigate the pixel intensity and contrast characteristics of the radiographs (please refer to Appendix B for full experimental details). The first involved exposing the whole CR plate and measuring the change in pixel intensity of the central part of the image with change in radiation dose incident upon the plate. This was used to determine the dose response of the CR system. The second method used a piece of aluminium milled as a series of steps (a so-called 'stepwedge') to determine the change in pixel intensities and contrast within any one radiograph as the radiation dose changed. Both this and the first method used the CR reader in raw mode[1], the mode in which little image processing is performed to map the pixel intensities to those required for clinical images. Using this mode, it was possible to test models of the relationship between both pixel intensity and noise, and the radiation dose incident on the plate. The final method used a specially constructed hand phantom and a series of radiographs to simulate how the pixel intensities and contrast change for different radiation doses to the plates when the plates were processed using the normal CR system 'hand' protocol - thus simulating

[1]The term 'raw mode' is used in this chapter to refer to the Kodak CR processing mode of 'Pattern' where there is minimal processing of the pixel intensities.

the clinical situation. This final method included placing a small sheet of aluminium over the objects in the carpal section of the phantom to simulate radiographic characteristics of overlapping bones.

4.2.1 Dose response of the Kodak CR system

In computed radiography a photostimulable phosphor plate is exposed to radiation that causes energy to be stored in a phosphor layer[2]. This stored energy is read by scanning the plate using a focussed laser beam that stimulates the emission of light of a shorter wavelength than the laser. This process is referred to as photostimulated luminescence. The light is collected using optics and a filter is used to separate the stimulating laser light from the emitted light. After detection and amplification of the filtered light signal, the resulting output signal is proportional to the radiation exposure of the plate at the point the laser is scanning. The output signal is linear with radiation exposure over a dynamic range of at least 10^4 [Rowl02]. This wide dynamic range cannot, however, be used to generate a meaningful radiographic image and the pixel intensities need to be mapped to values suitable for display of the image. The manufacturers of CR systems reduce the dynamic range using logarithmic or square-root processing that is implemented using either analogue or digital circuits. In the Kodak CR 900 system the detected light signal is converted to a digital value using a 16-bit analogue-to-digital converter. However, the output of the analogue-to-digital converter undergoes a digital 16-bit to 12-bit logarithmic conversion [Bogu95]. The resulting 12-bit signal undergoes correction for variation in the efficiency of the light collection optics, correction for variation in the scanning laser output power, and tone scaling to produce useful images irrespective of the radiation dose to the CR plate [Bogu95].

The dose response of the CR plate is sometimes referred to as the signal transfer property and is the relationship between the dose to the CR plate and the output pixel intensities. This dose response is usually stated for a particular kVp setting of the x-ray generator and for a particular filtration of the x-ray beam. For example, for the Kodak Ektascan HR plates this relationship is [Bogu95]

$$P = 1000 \log(D) + 1700, \tag{4.1}$$

where P is the intensity of the pixels in the image 15 minutes after exposing the plate[3], D is the CR plate entrance dose in micro-grays for an x-ray generator voltage of 80 kVp, and the x-ray beam is filtered with an additional 0.5 mm copper and 1 mm aluminium [Same01]. The common approach to analysing the dose response curves for CR systems is to use a logarithmic response to fit the data using the natural logarithm relationship $P = a \ln(D) + b$,

[2] Many important details about the operation and performance of CR technology and systems can be found in the publications by Rowlands [Rowl02] and AAPM Task Group 10 Report 93 [Seib06].

[3] Please note that this pixel intensity relationship is for the raw data in the CR system. The output pixel intensities are these values subtracted from 4095.

where a and b are constants [Mars09] [Mack08]. However, the relationship between $ln(D)$ and $log(D)$ is a simple scaling by $log(e)$; so this relationship is still of the form of Eqn. (4.1). Without accurate information about the Kodak CR 900 plate reader[4], it is reasonable to assume that because the Kodak CR system uses a logarithmic conversion built into the overall analogue-to-digital conversion [Bogu95], it is likely that this is implemented as a look-up table, and both the logarithmic operation and scaling are fixed. This would produce discrete steps in the output values. By using a method to reduce the influence of the corrections for the optics collection efficiency, these discrete steps have been observed by the author for CR images [OKee04]. Using this assumption, the dose response for the Kodak CR system can be modelled as

$$P = 1000 \log(D + c) + b, \tag{4.2}$$

where b and c are constants. Different x-ray tube kVp settings will produce different doses, but by modelling these doses as simple scaling of D by a factor k (and ignoring c for now), then the new dose will produce a pixel intensity of $P = 1000 \log(kD) + b = 1000 \log(D) + (1000 \log(k) + b)$. So the pixel intensities for different kVp settings simply result in changes to the constant added to the dose response relationship (the curves move up and down). Similarly, poor calibration of the dose measurement equipment will result in a scaling of the dose, and a similar additive shift in the model. What is important to note is the same applies to any factor that results in an apparent linear scaling of D, for example, scaling of the light detection signal.

This logarithmic relationship is a simplified model of the CR plate dose response because when the dose is zero the model should theoretically produce values approaching negative infinity. The values typically do not even go to zero because of the inability to completely erase the CR plate after reading it and because of dark currents in the light detection circuitry. Both of these result in a lower, positive-valued limit to the pixel intensities. The dose response models do not allow for this. In an attempt to allow for offsets, incomplete erasure of the CR plate, and possible offsets in the dose measurements, the factor c was included in the model Eqn. (4.2). The validity of this model is demonstrated in Figure 4.1 for radiographic exposures of 52 kVp and 70 kVp. An alternative logarithmic plot of the same data and model is presented in Figure 4.2 where the performance of the models is more clearly illustrated for low doses. There was difficulty in producing controllable radiographic exposures for low doses at 52 kVp; it was easier to control the 70 kVp doses because of the added x-ray beam filtration. This meant that the doses entering the CR plate were truncated at a level higher than in some clinical radiographs because in clinical radiographs the hand attenuates the x-ray beam and reduces the subsequent dose to the CR plate. A future improvement to this experiment would be to place a sheet of perspex in

[4]The author attempted to secure this information from the manufacturer, but was unsuccessful.

Table 4.1 Dose response of the Kodak Ektascan HR plates and Kodak CR 900 plate reader at 52 kVp and 70 kVp: parameters derived from fitting different models to the CR plate dose response. P is the image pixel intensity, D is the entrance dose to the CR plate in μGy, and a, b, and c are derived constant parameters. R^2 is the correlation coefficient for the fit. The results illustrate the improved performance of the models when only clinically relevant doses of less than 20 μGy are analysed. Uncertainties are the standard error for the parameter.

Model	a	b	c	R^2
52 kVp				
$P = 1000\log(D+c)+b$	1000	1335±13	-0.22±0.16	0.99484
$P = a\log(D)+b$	996±20	1329±25	-	0.99429
$P = 1000\log(D+c)+b$ fitted over a limited dose range of D<20 μGy.	1000	1385±4.5	-0.62±0.06	0.99950
$P = a\log(D)+b$ fitted over a limited dose range of D<20 μGy.	1107±17	1244±16	-	0.99800
70 kVp				
$P = a\log(D)+b$	1057±46	1534±46	-	0.98862
$P = 1000\log(D+c)+b$ fitted over a limited dose range of D<20 μGy.	1000	1660 ±14	-0.19±0.02	0.99862

front of the CR plate and dose meter, to allow measurements at even lower doses.

Initial fitting of the model of Eqn. (4.2) to the dose response data produced a good fit for low doses, but relatively large departures for high doses. This was the case for both model Eqn. (4.2) and a more general model of the form $P = a\log(D)+b$. However, for clinical hand-wrist radiographs at the typical x-ray tube setting of 52 kVp, hand entrance doses of approximately 40 μGy will result in radiation doses entering the plate of less than 20 μGy for soft tissue thicknesses of more than 20 mm. Only using doses less than 20 μGy improved the fit for low doses – the doses that were clinically relevant. The same result was observed for 70 kVp measurements. The derived parameters for the different models are listed in Table 4.1. It is evident from the results that the model works well for the CR entrance doses of interest (D<20 μGy), but the values of $c < 0$ cannot be explained physically and are a likely limitation of the data collected for low doses, or an incorrect model of the low dose response of the CR system.

The CR system manufacturers map the CR output values to something similar to those of a screen-film x-ray system. The need for this is illustrated in Figure 4.3 where a CR plate dose response is compared with a typical screen-film system for hand-wrist radiographs. The screen-film system has a section with an essentially linear relationship between the logarithm of the entrance dose and the film blackening (optical density). This section of the screen-film curve has relatively high contrast, and the response of the CR system has

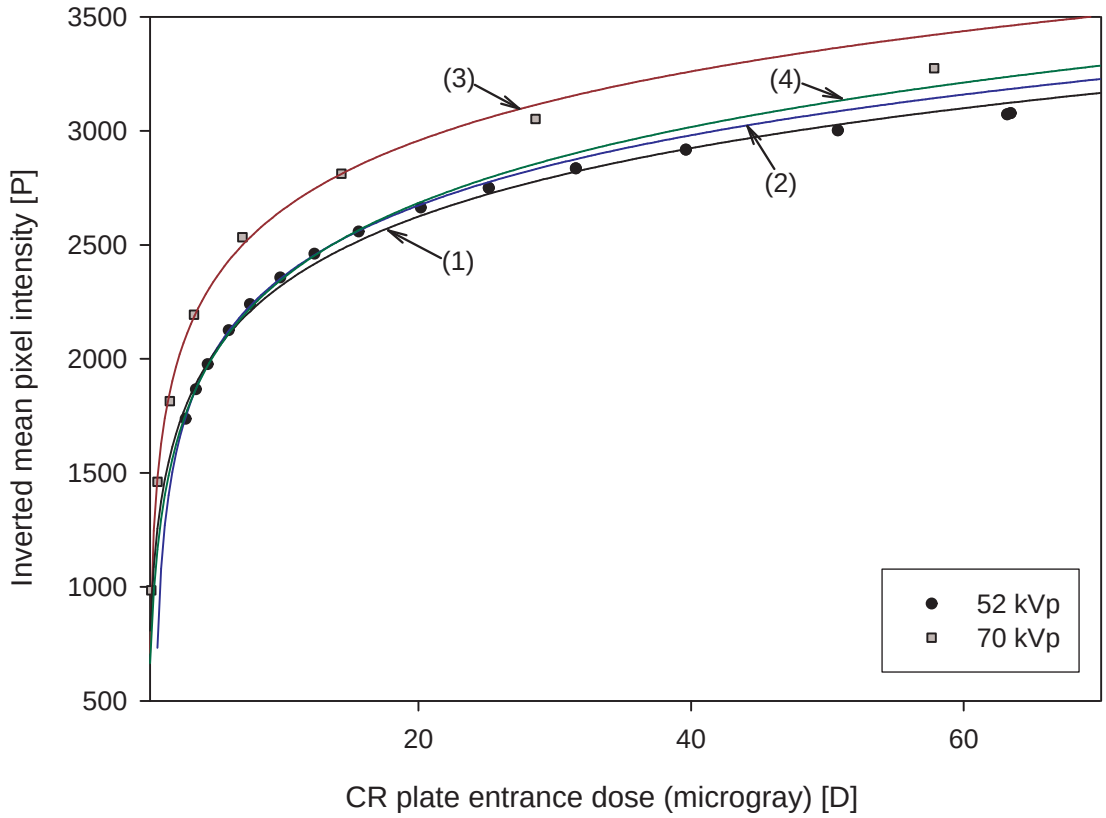


Figure 4.1 Dose response relationship for the Kodak Ektascan HR plates and CR 900 plate reader. The exposures were at 52 kVp and 70 kVp and pixel intensities (P) have been derived from 4095-(CR plate pixel value): (1) the generalised model of $P = a\log(D) + b$ for 52 kVp (the curve was very similar for the model $P = 1000\log(D + c) + b$ over the whole range); (2) the model $P = 1000\log(D+c)+b$ for 52 kVp and fitting using values of D<20 μGy; (3) model $P = 1000\log(D+c)+b$ for 70 kVp and fitting using values of D<20 μGy; and (4) the generalised model of $P = a\log(D) + b$ and fitting using values of D<20 μGy.

to be modified to match this contrast. Furthermore, unlike screen-film which has a fixed curve and the image is only visible if a certain optical density has been exceeded, the CR response can be remapped to different output pixel intensities at any dose. This is one of the major advantages of the CR technology in that it is possible to use a range of x-ray exposures to produce similar looking images.

4.2.2 Pixel intensity characteristics

The Kodak CR plate reader uses a series of image processing algorithms to produce radiographs that look similar for different radiation doses to the CR plate. These algorithms are applied as part of the clinical processing protocol associated with both the selected body part being examined and radiographic projection for the body part. In this case the 'Hand' body part and 'PA' projection (palm-down on the x-ray cassette) were used locally

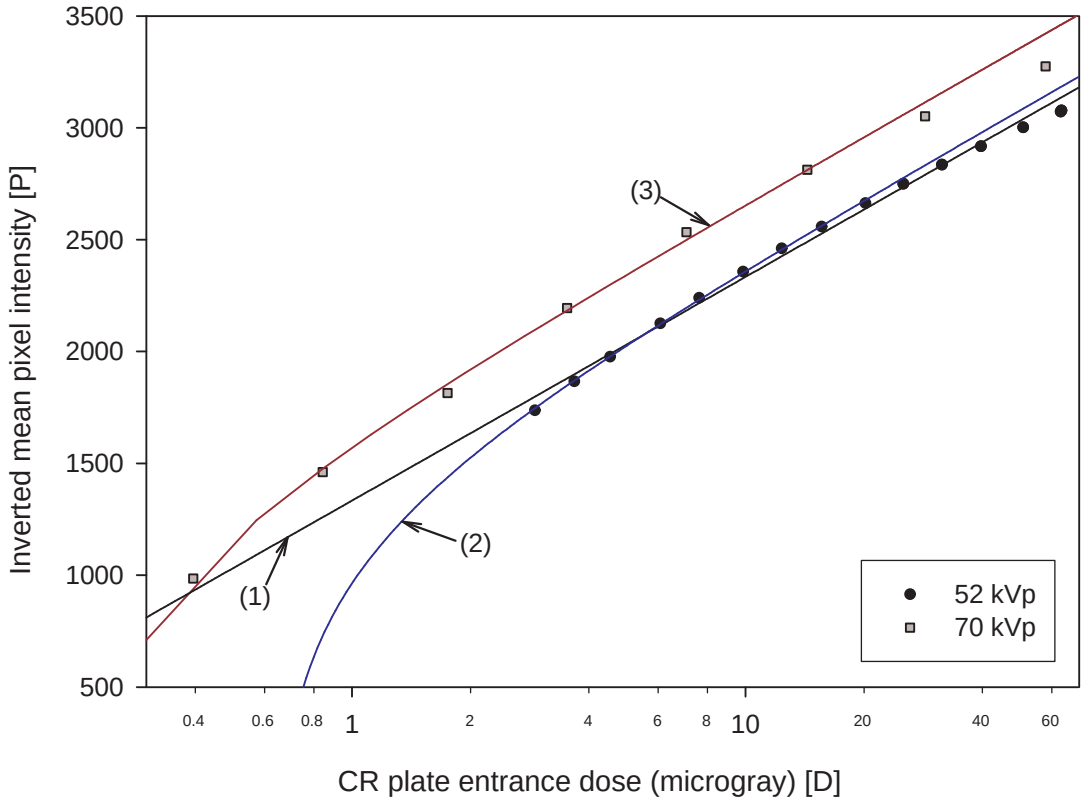


Figure 4.2 An alternative presentation of the results presented in Figure 4.1 for the dose response relationship for the Kodak Ektascan HR plates and CR 900 plate reader. The exposures were at 52 kVp and 70 kVp and pixel intensities (P) have been derived from 4095-(CR plate pixel value). Logarithmic scale for dose: (1) the generalised model of $P = a\log(D) + b$ for 52 kVp, fitted across the whole range; (2) the model $P = 1000\log(D + c) + b$ for 52 kVp and fitting using values of D<20 μGy; (3) the model $P = 1000\log(D + c) + b$ for 70 kVp and fitting using values of D<20 μGy.

for hand-wrist images for bone age assessment.

When the 'Hand' clinical processing protocol is used the Kodak CR system processes the radiograph to find the radiographic field (the main exposed area) and uses segmentation to find the hand outline so that only pixels of clinical significance are used to determine the processing parameters [Bogu95]. The finding of pixel values of interest is followed by tone scaling and visual enhancement such as edge enhancement. The combined processing meant that it was not possible to investigate the pixel intensity and contrast characteristics of hand radiographs without using something that looked radiographically similar to a hand. Because of the difficulty in justifying multiple radiographs of children to investigate pixel intensities, a hand phantom was developed to radiographically simulate a hand (please refer to Appendix B for details). This hand phantom contained regions of uniform attenuation with no structure so that noise and contrast measurements could be made.

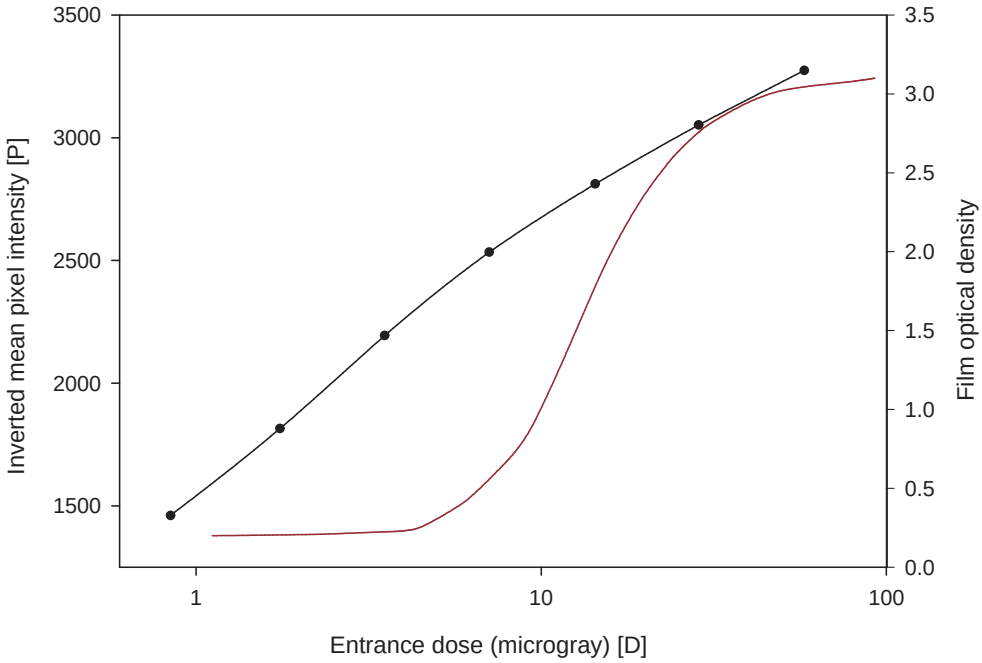


Figure 4.3 The dose response of the Kodak CR system (•) and the dose response of a typical 100-speed screen-film system for hand-wrist radiography (no markers). To produce images that look similar to the optical density curve of the screen-film system, the output pixel intensities of the CR system need to be adjusted to give a dose response curve similar to the screen-film system.

Details of the phantom construction are in Appendix C.

The hand phantom was x-rayed over a wide range of doses. The seven regions of interest within the phantom for the mean pixel intensity measurements are shown in Figure 4.4. These seven regions correspond to different amounts of attenuation and, hence, different doses to the CR plate below the corresponding part of the phantom. In these measurements the entrance dose refers to the entrance dose to the hand phantom or test object, not the actual radiation dose incident upon the CR plate. For the local hand-wrist radiographs, this entrance dose to the phantom was approximately 40 μGy. Just below this dose the mean pixel intensities of the regions of interest increase sharply. This is thought to be caused by the Kodak CR plate reader implementing a low-dose optimisation method for controlling the image noise – this function was left on for these radiographs, as per the clinical protocol. As shown in the previous section, the usual variation of pixel intensity with dose is a logarithmic relationship. The tone scaling algorithm in the CR plate reader works well to maintain the mean pixel intensities across a wide range of doses, although there are changes that occur for low doses. The limitation of this experiment is that the geometry of the hand phantom is fixed and this does not test the change in pixel intensities with, for example, collimation of the image. Such assessment would require construction

of hand phantoms with multiple hand sizes and this has been reserved for future research. (Collimation of the image could cause truncation of the simulated soft tissue of the hand, possibly affecting the scaling.)

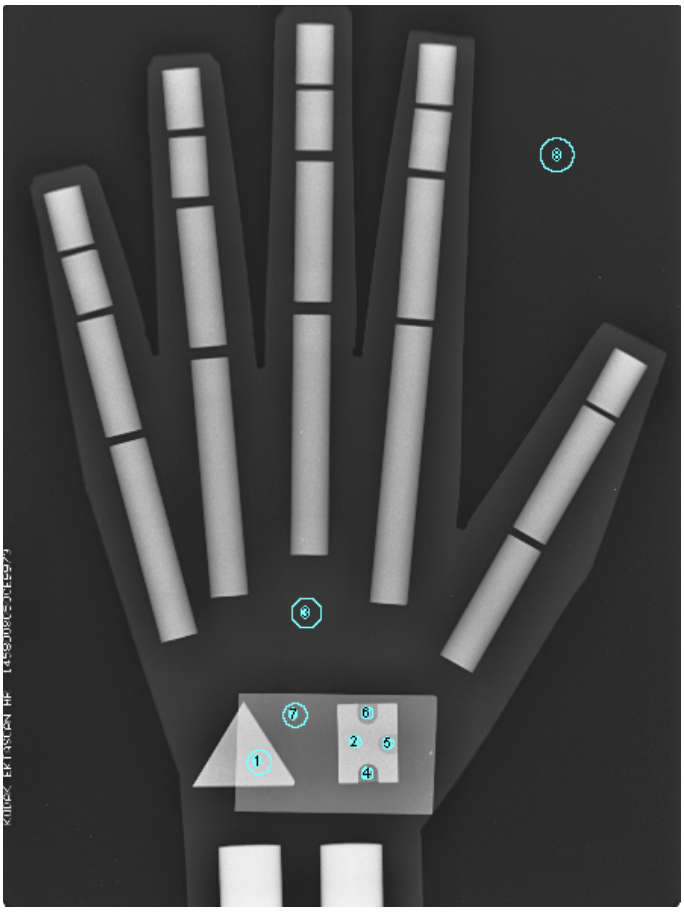


Figure 4.4 Regions of interest used for measurement of the pixel intensities and noise of the hand phantom radiographs. This radiograph shows the aluminium plate used to simulate overlapping carpal bones.

4.2.3 CR contrast changes with dose

Although the absolute pixel intensities are of interest for image processing, equally important for both computer vision and clinical interpretation is the contrast within the image. The contrast change with dose for the Kodak CR system was measured for the raw image mode using an aluminium stepwedge, and for the clinical mode (Hand) using a hand simulating phantom. The contrast C_{ab} between two regions a and b was calculated using the formula

$$C_{ab} = \frac{|I_a - I_b|}{\frac{1}{2}|(I_a + I_b)|}$$

where I is the mean pixel intensity in a region.

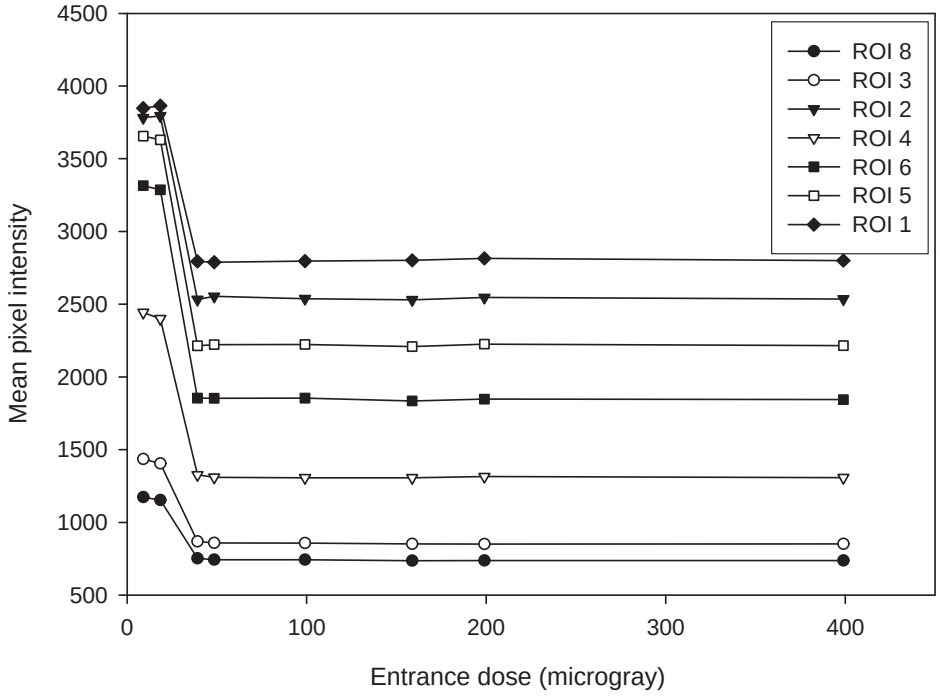


Figure 4.5 Mean pixel intensities for seven regions within the hand phantom. The variation in pixel intensity with dose would normally be logarithmic. The flat lines illustrate the ability of the CR plate reader to maintain pixel intensities and contrast with increasing dose. The typical dose for hand-wrist radiographs is approximately 40 μGy. ROI 7 was excluded from the analysis because it was only used for the simulation of overlapping bones and has essentially the same intensity as ROI 3 .

Radiographs of the aluminium stepwedge were processed using the Pattern raw mode, hence no tone scaling or visual enhancement was applied. The contrast between steps of equal thickness increased slightly with dose for steps that had high pixel intensities (corresponding to greater thickness of the wedge), but the increases with dose were largest for the low pixel intensity steps, as illustrated in Figure 4.6. Because the energy spectrum of the x-ray tube is approximately constant with increasing dose for modern x-ray systems – and this was confirmed for these experiments by measuring the kVp – the source of contrast variation for any single step is most likely the CR system not having a simple logarithmic dose response for higher x-ray exposures. As introduced earlier, if the dose is increased from D to kD, the pixel intensities increase by a constant factor of $log(k)$. This means that the contrast between two steps will not increase. However, for a departure from a logarithmic response for higher x-ray doses, the relationship $log(kD) = log(k)+log(D)$ no longer applies and the contrast between steps will change as the dose changes. This effect will be largest for steps that already produce a high exposure to the plate and this was observed in this experiment. The change in contrast between steps for any fixed dose can be explained by preferential attenuation of lower energy components of the x-ray beam

as it passes through the thickness of aluminium. This results in decreasing low energy components in the x-ray beam, so overall there is progressively less attenuation because the remaining higher energy components are attenuated less. This means that identical thickness steps will have different contrast depending on whether they are in thick or thin parts of the stepwedge.

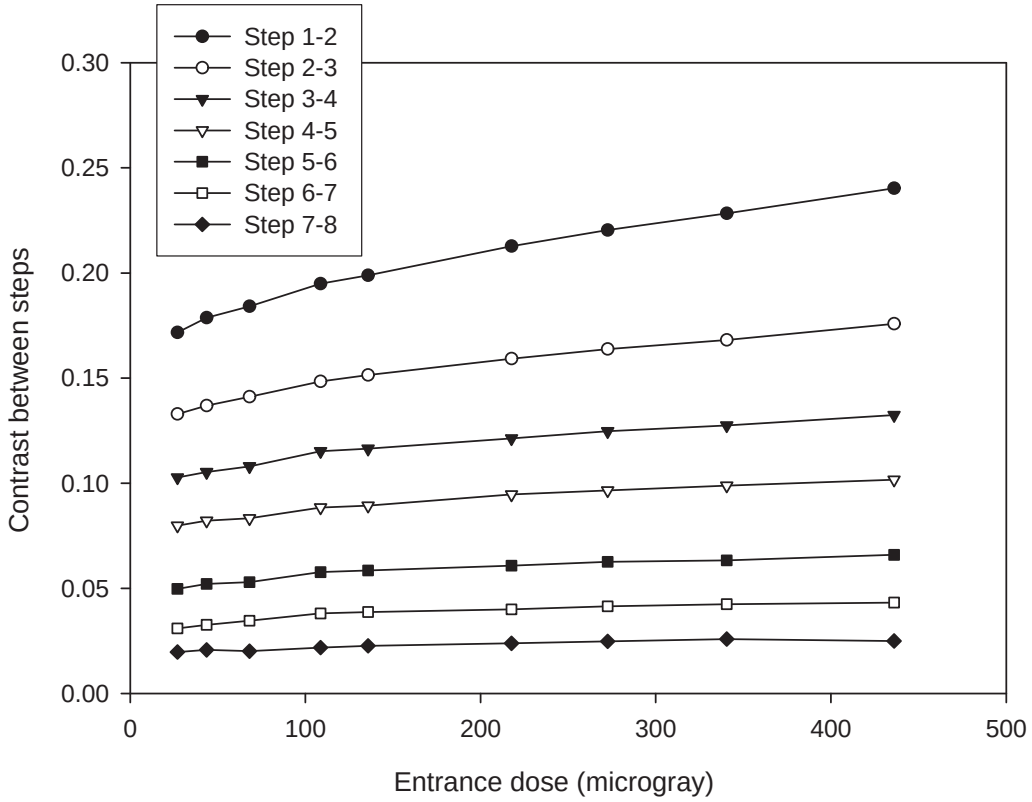


Figure 4.6 The change in contrast with dose for steps of an aluminium stepwedge processed using an operation mode that does not use tone scaling or visual enhancement algorithms. The contrast between the steps is not equal because the attenuation characteristics of the x-ray beam changes as the beam passes through thicker sections of aluminium (referred to as 'beam hardening'). For any one step, these attenuation characteristics should be fixed, so the curves should ideally be flat. Contrast increases with dose the most for low pixel intensities (lower step numbers) that correspond to areas of increased exposure of the radiograph. This indicates dose-dependent contrast changes.

The stepwedge images show a slight variation in contrast for the dark areas of the image, but the clinical processing protocol appear to maintain contrast in the hand phantom image across a range of doses, as illustrated by the maintenance of pixel intensities in Figure 4.5. This slight difference in performance could reflect the ability of the clinical processing algorithms to maintain image characteristics across a wide range of doses. The performance difference is unlikely to be caused by attenuation differences between the hand phantom and the stepwedge because the hand phantom contained a wide range of effective attenuators of the x-ray beam and sufficient low attenuation regions to produce similar dark regions of the image compared with the stepwedge.

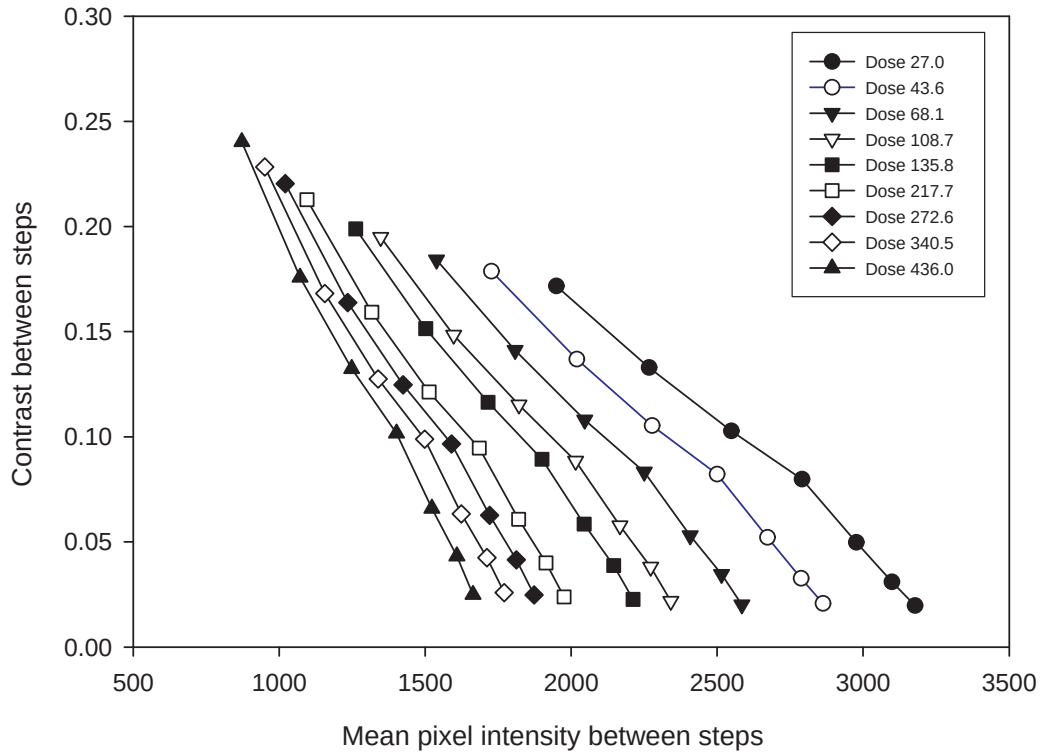


Figure 4.7 An alternative analysis of the results in Figure 4.6, clearly illustrating the increased contrast for lower pixel intensities, and the slight increase in contrast with increasing dose also for lower pixel intensities. The line for a typical entrance dose for clinical hand-wrist radiographs ($\sim 40\,\mu$Gy) is marked with open circles ($\circ$). All doses are in units of μGy.

4.2.4 The contrast of overlapping objects

One image contrast that is especially relevant to this research is the scenario of overlapping objects. When two thin objects overlap, the resulting fractional x-ray transmission should be the product of the transmission of each object. This means that when a logarithmic dose response is present, as in the Kodak CR system, the fractional transmission should theoretically be a simple sum of the logarithm of each individual transmission component (sum of the CR pixel intensities). However, this model of the combined transmission assumes that there is little scattered radiation and that the x-ray beam contains a single energy – this is not the case for the output from x-ray tubes. As seen earlier with the contrast for the steps of the aluminium stepwedge, the overlap of bones is unlikely to be a simple product of the transmissions or sum of the CR pixel intensities.

The effect of bone overlap on pixel intensities in clinical radiographs was tested by placing a 5 mm aluminium sheet over part of the carpal objects in the hand phantom. Rather than a fixed increase in pixel intensity, Figure 4.8 shows that the increase is linearly dependent on pixel intensity, and that it changes significantly at low doses when the low-dose

optimisation of the CR system activates. The results presented in Figure 4.7 suggest that the contrast increases approximately linearly as the mean pixel intensity of the stepwedge steps decreases. This supports the observation that for the overlap, the contrast increases as the pixel intensities decrease because in this situation the overlap is similar to the next step on the stepwedge. Furthermore, the very similar curves for the two higher doses of 48.5 and 199 μGy are in agreement with the maintenance of pixel intensities in clinical radiographs, as illustrated in Figure 4.5. These findings mean that there is no simple relationship between the pixel intensities for overlapping objects for both the raw mode and clinical mode of processing of the radiographs. Therefore, it is not possible to identify overlapping objects by relying upon the simply addition of pixel intensities of the individual objects: visually the overlap will appear to be the sum of the individual pixel intensities, but mathematically they are not a simple sum. This is important for the segmentation of overlapping carpal bones because it means that the overlap cannot be be identified by assuming, for example, that the overlap of two homogenous regions will result in a summation of the pixel intensities in the regions.

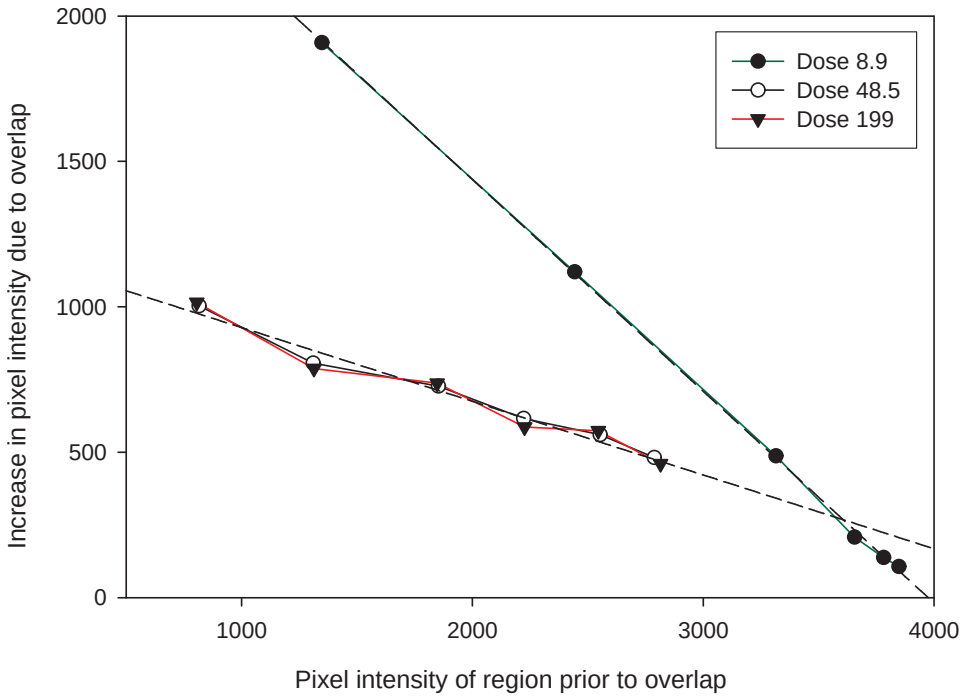


Figure 4.8 Simulation of bone overlap using the hand phantom and a sheet of aluminium. For simple attenuation the curves should be flat – independent of the pixel intensity. The significant difference for the low dose of 8.9 μGy can be explained by activation of the low-dose optimisation of the Kodak CR system. The dashed lines are linear regression fits. All doses are in units of μGy.

4.3 Radiographic noise

Computed radiography images contain noise caused by a number of sources. These include quantum fluctuations in the x-rays producing the image, conversion noise in the image receptor, light output fluctuations from the CR plate, electronic noise in the detection and processing of the light, quantisation noise in the digital conversion, and structural noise from spatial gain variations in the CR plate [Bogu95, p23]. Some of these are truely random noise sources, others, like structural noise, may be more systematic and simply appear as a noise component.

A number of comprehensive models of radiographic noise have been published (see, for example, [Barn82]). Many of these models are formulated in the spatial frequency domain because they address the interest in contrast-detail performance of the imaging systems. Although these models are useful for understanding the relationship between design characteristics and performance of the imaging systems, the models presently have unwarranted complexity for the task of modelling image noise for the purpose of denoising. The approach taken in this research was to develop and test a simple model of the noise, one designed to be workable for image denoising.

The relative contribution of the different noise processes in radiographs depends on the dose incident on the CR plate. The quantum noise of x-ray production is the dominant noise source in radiographs taken using typical clinical exposures [Mack07]. To produce a simple noise model, only quantum noise sources have been considered. Furthermore, it needs to be noted that even if all the quantum noise can be removed from an image, the image will still contain a type of textural noise due to the bone structure. Unfortunately, this bone structure varies significantly from person to person and is not consistent enough to be used for texture-based segmentation. So at this stage, investigation and removal of other minor noise sources is likely to have a diminishing return because of the significance of this textural noise.

Although not strictly a noise source, superimposed on the image structural components and quantum noise is scattered radiation detected as part of the radiographic process. The high density and high effective atomic number of bones in the hand means that the main x-ray interaction is through the non-scattering photoelectric effect [John83]. This is in contrast with soft tissues of the hand in which compton scattered radiation is the main interaction of the x-rays. Although the scattered radiation from the soft tissue reduces observed image contrast, Monte Carlo simulations have shown that this scattered radiation has only low spatial frequency components [Wins05]. This means that there will still be noise due to the detected scattered radiation, but the scattered radiation will not directly contribute high spatial frequency components to the image.

The most common approach to modelling image noise across many applications is to assume it is Gaussian white noise with a fixed variance. This model is not valid in many radiographic images because the quantum fluctuations in the number of x-rays generated follows a Poisson statistic.

4.3.1 A Poisson noise model

An x-ray tube can be treated as a Poisson source and fluctuations in the number of x-rays emitted from the tube can be described by a Poisson probability density function [Barr04, p1103]. Many of the detection processes for the x-rays can be described by either binomial selection or Poisson processes and it can be shown that the cascading of these processes with the Poisson x-ray tube source results in another Poisson distribution [Barr04, p638] [Yaff94].

This simplified model of a noise source with a single Poisson distribution can be used to derive a theoretical model of the relationship between the image pixel intensities and the noise in these pixels. The noise measurement of interest is the variance of pixel intensities in a small region of interest within the image. If measurements are made using regions of interest in the image that have a large area compared with the blur of the detector, then pixel intensities can be considered independent and identically distributed, and the mean pixel intensity of the region is the mean of a collection of uncorrelated pixels [Barr04, p1101]. This independence means that the noise can be considered individually for each pixel in the region. The region of interest has to be small compared with gross variations in the x-ray flux across the detector. This is so that there is no variation across the sampling region and the expected value ($\mathrm{E}\{.\}$) of the individual pixels in the region equates to the mean pixel intensity in the region. This requirement also means that the measured variance of the pixel intensities in the region is only due to the random noise and not the gross x-ray flux variations.

As introduced in Section 4.2.1, the output from the CR plate light detection system is mapped to pixel intensities by a logarithmic calculation of the form

$$Y = a\log(X) + b \tag{4.3}$$

where a = 1000 and b incorporates scaling of X via the relationship $log(kX) = log(k) + log(X)$. If X and Y are treated as random variables, then the mean and standard deviation for a measurement region of interest can be calculated by applying a probability density function based on a Poisson distribution, but with a change of random variables based on Eqn. (4.3) [Barr04, p1443] [Pres07, p362].

The proposed noise model is a single Poisson distribution. A discrete random variable X

is said to be Poisson distributed or Poissonian if for a parameter λ it obeys the probability law

$$p(X = x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

where x is a non-negative integer for the number of events in an interval. The parameter λ is the expected number of occurences in an interval. The expected value (mean) μ and variance σ^2 of X are both equal to λ [Barr04, p1467], so the probability law for the Poisson distribution can be written

$$p(X = x) = \frac{\mu^x e^{-\mu}}{x!}$$

meaning that the random variable X is completely specified by its mean value μ. If $\mu > 10$ then the discrete Poisson distribution can be approximated by a continuous normal distribution with mean μ and variance μ [Barr04, p1468]:

$$p(X = x) = \frac{1}{\sqrt{2\pi\mu}} e^{-\frac{(x-\mu)^2}{2\mu}}. \tag{4.4}$$

4.3.1.1 Mean pixel intensity of a noise region of interest

If X is a random variable for the x-rays detected in the measurement region, then the mean, P, of the pixel intensities is given by the expectation operation $P = \mathrm{E}\{Y\} = \mathrm{E}\{a\log(X) + b\}$.

Using the linearity of the expectation operation, $\mathrm{E}\{Y\}$ can be simplified to $\mathrm{E}\{a\log(X) + b\} = a\,\mathrm{E}\{\log(X)\} + b$. The expectation $\mathrm{E}\{\log(X)\}$ can be approximated using a Taylor series expansion and only keeping terms to a second order [Barr04, p1445]:

$$\begin{aligned}
\mathrm{E}\{\log(X)\} &= \log(\mathrm{E}\{X\}) + \frac{1}{2}\sigma_X^2 \left.\frac{d^2\log(x)}{dx^2}\right|_{x=\mathrm{E}\{X\}} \\
&= \log(\mathrm{E}\{X\}) - \frac{1}{2}\sigma_X^2 \frac{1}{ln(10)\,\mathrm{E}\{X\}^2} \\
&= \log(\mathrm{E}\{X\}) - \frac{1}{2}\frac{1}{ln(10)\,\mathrm{E}\{X\}}
\end{aligned}$$

because for the random Poisson variable $X, \sigma_X^2 = \mathrm{E}\{X\}$.

So for $\mathrm{E}\{Y\} = \mathrm{E}\{a\log(X) + b\}$,

$$\begin{aligned} P = \mathrm{E}\{Y\} &= a\,\mathrm{E}\{\log(X)\} + b \\ &\approx a(\log(\mathrm{E}\{X\}) - \frac{1}{2}\frac{1}{\ln(10)\,\mathrm{E}\{X\}}) + b \\ &\approx a(\log(\mathrm{E}\{X\}) - \frac{0.2171}{\mathrm{E}\{X\}}) + b \\ &\approx a\log(\mathrm{E}\{X\}) + b. \end{aligned} \tag{4.5}$$

As the expected value of X increases, this approximation improves, although $\mathrm{E}\{\log(X)\}$ remains slightly less than $\log(\mathrm{E}\{X\})$. This is supported by the result that because $\log(x)$ is a concave function, the opposite of Jensen's inequality for the relationship between functions of random variables and expectations [Barr04, p1444] can be applied to give $\mathrm{E}\{\log(X)\} \leq \log(\mathrm{E}\{X\})$ (recalling that $X > 0$ must apply for physical detection of an x-ray). Furthermore, the approximation of a Poisson distribution with a normal distribution only applies for $\mathrm{E}\{X\} > 10$, and for these values the error in the approximation of $\mathrm{E}\{Y\}$ is less then 2.2%. As the CR pixel intensities (P_{inv}) increase, the value of $\mathrm{E}\{X\}$ decreases (because $P_{inv} = 4095 - \mathrm{E}\{Y\}$) and the measurement of the mean becomes more inaccurate. This is an important caveat when analysing areas of high attenuation in the images, for example, thick bones.

4.3.1.2 Variance of a noise region of interest

For a sampled region of interest, the variance of the random variable Y is a measure of the image noise. A first order approximation to the variance is given by [Jahn02, p82] [Knol00, p87]

$$\mathrm{Var}\{Y\} = \mathrm{Var}\{X\}\left|\frac{dy}{dx}\right|^2_{x=\mathrm{E}\{X\}}$$

Using $y = a\log(x) + b$ and because for a Poisson distribution of X, $\mathrm{Var}\{X\} = \mathrm{E}\{X\} > 0$

$$\begin{aligned} \mathrm{Var}\{Y\} &= \mathrm{E}\{X\}\left|\frac{a}{\ln(10)\,\mathrm{E}\{X\}}\right|^2 \\ &= \frac{a^2}{\ln(10)^2\,\mathrm{E}\{X\}} \end{aligned}$$

Applying the approximation $\mathrm{E}\{X\} = 10^{(\frac{\mathrm{E}\{Y\}-b}{a})}$ (inverting Eqn. (4.5)), this equation can be written in terms of the variable Y only

$$\begin{aligned} \mathrm{Var}\{Y\} &= \frac{a^2}{\ln(10)^2}10^{-(\frac{\mathrm{E}\{Y\}-b}{a})} \\ &= \frac{a^2}{\ln(10)^2}10^{(\frac{P_{inv}+b-4095}{a})} \end{aligned} \tag{4.6}$$

where the mean pixel intensity substitution $P_{inv} = 4095 - P = 4095 - \mathrm{E}\{Y\}$ has been applied.[5]

Taking the square root of the variance to give the standard deviation (σ) and taking the logarithm of both sides gives

$$\begin{aligned} \log(\sigma) &= \log(\frac{a}{\ln(10)}) + \frac{1}{2}\left(\frac{P_{inv} + (b - 4095)}{a}\right) \\ &= \left(\log(\frac{a}{\ln(10)}) + (\frac{b - 4095}{2a})\right) + \left(\frac{P_{inv}}{2a}\right) \end{aligned} \tag{4.7}$$

This noise model suggests that for radiographs in which the x-ray statistics can be described by a single Poisson distribution (the quantum fluctuations dominate) and the transformation of pixel intensities is of the form of Eqn. (4.3), the image noise is signal-dependent and should vary approximately exponentially with the mean pixel intensity (P_{inv}). This model is very similar in form to the result stated by Gravel *et al.* of the exponential relationship between noise variance and optical density in screen-film radiography [Grav04]. Plotting the logarithm of the standard deviation for a region of interest versus the mean pixel intensity should produce a straight line with a slope of $\frac{1}{2a}$. (For the Kodak CR system $a = 1000$ – refer to Eqn. (4.2)).

4.3.2 Testing the noise model using simulation of Poisson noise

The expectation approximation introduced in the CR noise model of Eqn. (4.6) was tested by simulating the random variable X using a Poisson random number generator [Mart02, p102]. To simulate physical measurements of variance and mean pixel intensity, 100 x 100 pixel regions of interest (10 000 samples) were simulated for 40 values of λ for the Poisson distribution uniformly sampled from $\lambda \in [17, 667]$. The CR simulation was performed using $a = 1000$ and $b =$ in the noise model. As illustrated in Figure 4.9, there was a linear relationship between the logarithm of the standard deviation of the pixel intensities and the mean pixel intensities, supporting the noise model and the validity of the approximations used in Eqn. (4.6) for the values of λ simulated.

4.3.3 Testing the noise model using a stepwedge

The results from the stepwedge measurements were used to validate the noise model for the Kodak CR system when operated in the raw mode (minimal image processing). The stepwedge contained too few steps to test the model across a range of pixel intensities, so results from different measurements at different doses were used to produce a compos-

[5] Introduction of the term P_{inv} and this correction are necessary because of the radiographic convention of 0 being black and 4095 being white in the Kodak CR system.

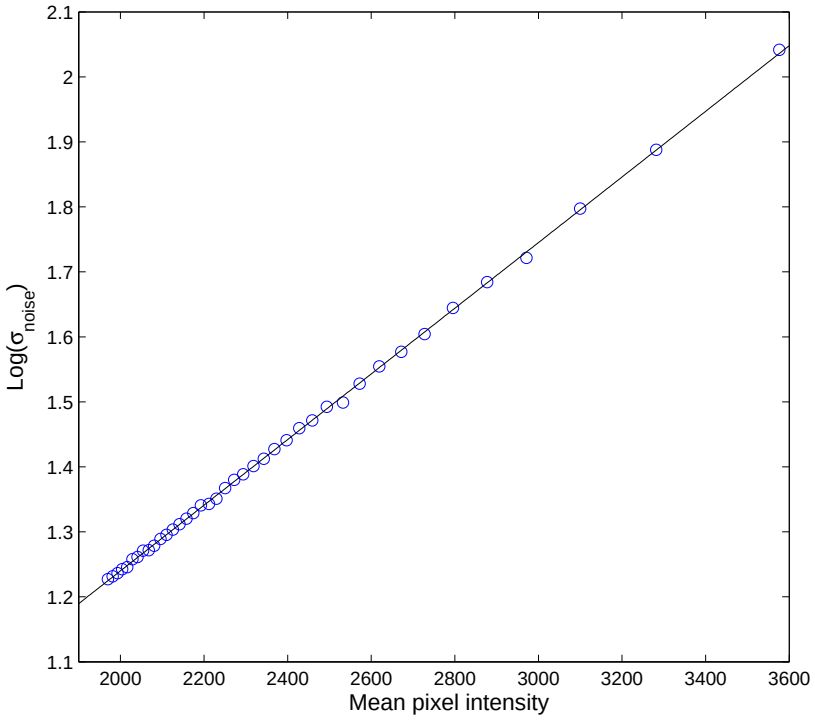


Figure 4.9 Simulation of Poisson distributed noise to test the approximations of the model from Eqn. (4.6). The data points are the logarithm of the standard deviation of 10 000 Poisson random numbers generated and transformed according to the Kodak CR system dose response model of Eqn. (4.3). The linear fit validates the noise model of the standard deviation of the noise increasing exponentially with the mean pixel intensity.

ite set of data to test the model. The noise versus pixel intensity results are illustrated in Figure 4.10. The linearity for low doses supports the proposed noise model for use in clinical radiographs, and fitting the linear model of Eqn. (4.7) with $a = 1000$ and the intercept as a free parameter of the fit produced a correlation coefficient of R^2 = 0.981 for pixel intensities larger than 1500. Pixel intensities less than 1500 correspond to the point at which the dose response curve presented in Figure 4.1 departs from the logarithmic model ($4095 - 1500 = 2595$ inverted pixel intensity in Figure 4.1). This is also the largest mean pixel intensity recorded for a set of 97 clinical radiographs before processing of the radiographs with a clinical processing protocol[6]. The results are obviously nonlinear for the low pixel intensities (larger doses), although it appears that the data may have two linear components, or at least all of the steps of the stepwedge for the higher dose radiographs are linear, but with a different slope than predicted by the model. One possible explanation for this departure from linearity is that the dose response relationship for the CR system

[6]This is the so-called Exposure Index (EI), a measure of the average pixel intensity in the anatomically important areas of the radiograph, in this case, the soft tissue and bones on the hand-wrist radiogaph. This is discussed further in Section 6.3.1.

used in these experiments did not follow the modelled logarithmic response at high doses. This is supported by the dose response results illustrated in Figure 4.1: the dose response 'rolls-off' faster than a modelled logarithmic response for high doses, and this results in a further decrease in the standard deviation of the noise. If the 'roll-off' of the dose response for high doses was a product of two logarithms, this would explain the possible appearance of mixed linear components in the data of Figure 4.10.

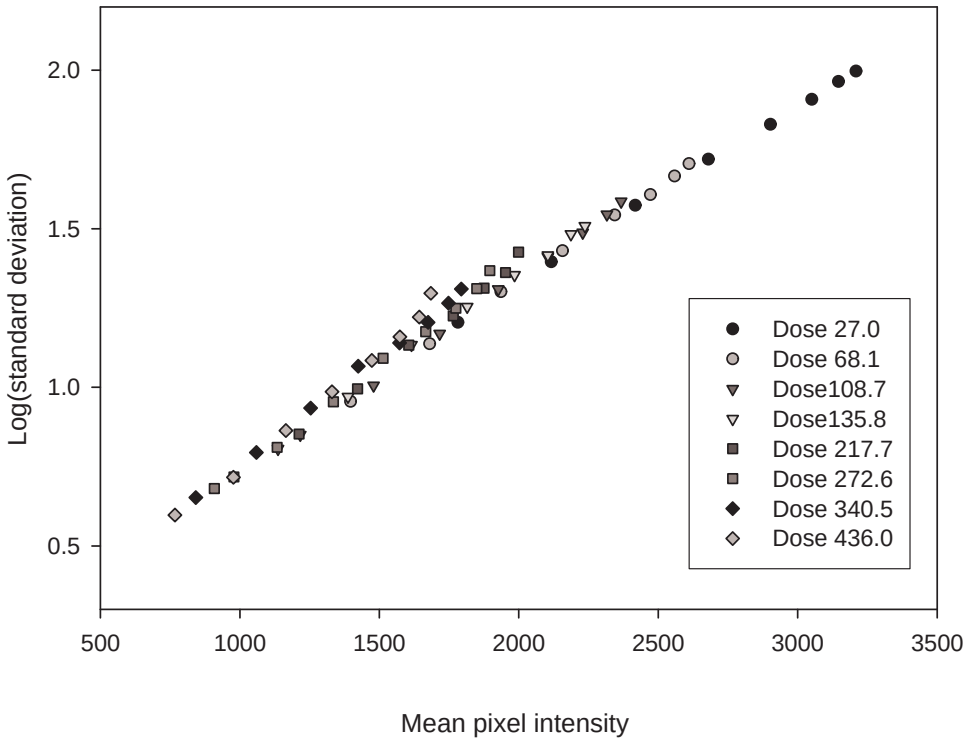


Figure 4.10 Kodak CR system radiographic noise for fixed regions of stepwedge images. This data is a composite of noise values from eight steps of the stepwedge, for eight different doses to the same plate ranging from 27 μGy to 436 μGy (the indicated doses in the legend are in units of μGy). The results for each exposure are labelled with different symbols. The pixel intensities are for minimal processing of the Kodak CR data (raw mode of processing). The pixels intensities of interest are those larger than 1500 and for these intensities the relationship is approximately linear with a slope in agreement with Eqn. (4.7) for $a = 1000$. However, as the dose to the plate increases the noise reduces faster than the model predicts and it appears that the data may have two linear components, one for low dose, and one for high doses.

4.3.4 Testing of the noise model using the hand phantom

In the previous section the noise model was shown to be valid for the unprocessed radiographs of a stepwedge at doses typical of those used for hand-wrist radiographs. The hand phantom was used to test this model for radiographs processed using clinical protocols. The hand phantom was intentionally designed with uniform attenuation areas that could

be used to measure the variance of the noise without structural features superimposed upon the image data. Regions of interest were used within these uniform areas to measure the variation of image noise with dose to the CR plate. To reduce variability, the same size and position of regions of interest were used, and the CR plate was positioned within a template to reduce changes in the position of the regions. There were seven useful regions within the hand phantom image, although one (region 4) produced erroneous results because it was of very large radiographic contrast, and its small size meant that the region of interest data was affected by edge enhancement routines used as part of the clinical image processing protocol. The results are presented in Figure 4.11 for the standard deviation on a log scale. At low doses the system applies low-dose compensation and this significantly changes the noise characteristics of the images, as observed for the 9 and 18.5 μGy curves. The logarithm of the noise is approximately linear for the clinically relevant doses of 39 and 48.5 μGy. Interesting to note is the sharp drop in noise for the the smallest pixel intensities. This region of interest is in the image background and these pixel intensities are normally not included in the calculations for the tone scaling of the image [Bogu95]. The drop could be due to a true reduction in noise in the background, or a compression of the image pixel intensity scale for the background region; that is, the pixel intensity for the background is higher than predicted from the standard deviation of the noise, so the data point is shifted by the mean pixel intensity.

Given the tone scaling and other image enhancements performed by the Kodak CR system, the results presented in Figure 4.11 for the clinically relevant doses of 39 and 48.5 μGy are a little surprising. Although these results validate the noise model of Eqn. (4.6) for radiographs processed using clinical protocols, the uniformity of the regions of interest used in the analysis would not normally be found in clinical radiographs. However, the results suggest that if regions of interest can be found that provide a reliable estimate of the image statistics in clinical radiographs, then it should be possible to measure the noise characteristics and fit a linear model to the logarithm of the standard deviation of the noise.

The results for higher entrance doses show a departure from linearity that can be explained by the non-logarithmic dose response of the CR system at these doses. This non-logarithmic response could result in a decreased standard deviation of the noise. Another explanation is the noise reduction due to the effect of the tone scaling for higher raw pixel intensities. Either way, such high doses are normally not used clinically, and hand doses as high as 99 μGy can still be reasonably well approximated with an exponential function (linear fit in Figure 4.11) if the dark areas of the image are avoided. These dark areas of the image are outside of the main region of interest of the hand and are not of importance once the hand outline has been detected for further processing.

The model of radiographic noise developed in this chapter for the Kodak CR systems is

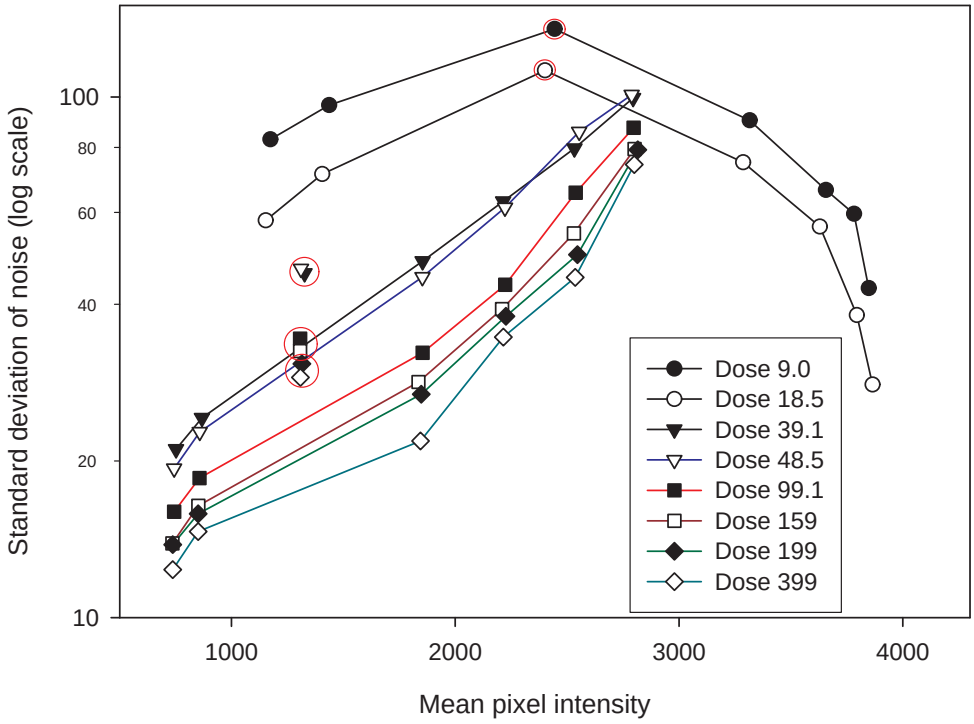


Figure 4.11 Kodak CR system radiographic noise for clinical processing of hand phantom images. The curves for the clinically relevant doses of 39 and 49 μGy support the proposed noise model being exponential with pixel intensity. The non-linear nature of the curves for doses over 49 μGy can be explained by the departure of the CR system from a logarithmic dose response for high doses, as also observed in Figure 4.10. The encircled data points are outliers corresponding to erroneous results for region of interest 4 of the images (refer to Figure 4.4). All doses are in units of μGy.

important for image denoising because it not only shows that the noise is non-stationary, it also indicates that there is significant variation in the amount of noise for the different pixel intensities in the radiographs. This can be seen from the results for the hand-phantom presented in Figure 4.11 where the variation in image noise (standard deviation) across the image is as high as a factor of six. If a denoising method uses an average value for the image noise, then some parts of the image will undergo excessive denoising, while other parts of the image will undergo insufficient denoising. This is especially important for the bones in the radiographs because these have the highest noise levels and require the most denoising. This is explored further in Chapter 6.

4.4 Conclusions

The Kodak computed radiography system uses a logarithmic transformation to convert the stimulated light output from the CR plate to image pixel intensities. Using this transforma-

tion and modelling of the radiographic noise as a single Poisson distribution, it was shown that the standard deviation of the noise is signal-dependent and varies approximately exponentially with increasing pixel intensities for radiographs processed using the raw mode on the CR plate reader. When clinical protocols are used to process simulated hand radiographs, the algorithms for tone scaling and visual enhancement maintain pixel intensities over a wide range of doses. This result is used in Chapter 5 for identifying image artefacts.

Even after tone scaling and visual enhancement, the image noise for uniform regions of the radiograph still has an approximately exponential relationship between the standard deviation of the noise and the pixel intensities. This model is used in Chapter 6.3 for noise estimation as part of image denoising.

Chapter 5

Preprocessing and region of interest extraction

The analysis of most images is performed in a series of stages. One of the first stages in the analysis of hand-wrist images is preprocessing. This stage is required to improve the accuracy and robustness of the image processing. In this chapter the preprocessing tasks of artefact identification and hand outlining are addressed. This preprocessing improves noise estimation methods (see Section 6.3), and creates an outline of the hand to initialise methods for region of interest extraction. Significant research has already occurred on the topic of region of interest extraction and this part of a bone age assessment system is essentially a development task. Rather than duplicate or parallel this research, the author chose to select the carpal region manually for this proof-of-concept research. This approach had the additional advantage of removing the region of interest extraction as a source of variability in the automated assessment of the carpal bones.

5.1 Introduction

Preprocessing and region of interest (ROI) extraction refer to image processing tasks that prepare the hand-wrist radiograph for bone segmentation and subsequent feature extraction for bone age estimation. The requirements for preprocessing depend on both the means by which the image was acquired and the subsequent image processing methods. For example, some of the computerised methods that worked with digitised screen-film radiographs had to be modified to work with computed radiography images [Piet01a]. The preprocessing takes an image considered suitable for reporting by a radiologist and changes it to a computerised representation that is suitable for segmentation and feature extraction. This means potentially changing the image characteristics like spatial resolu-

tion, contrast and noise. An important question is how much preprocessing is tolerable, that is, how much can the images change before these changes affect the computerised bone age measurements? This question has not been answered in the published literature. Similarly, and perhaps more surprising, there is no consensus on the image quality requirements for the manual methods of bone age assessment that are in use. One possible explanation for the lack of consensus may be the different methods of bone age assessment. For example, a high resolution screen-film combination is recommended for use with the TW3 method [Tann01], whereas it has been claimed that if a bone age precision of only 0.5 years is considered acceptable, then a reasonable estimate can be achieved using a much lower resolution bone density scanner combined with an atlas-based method similar to that of Greulich and Pyle [Plud04]. Because of the uncertainty regarding the image quality requirements, the approach taken in this research was to preserve spatial resolution and image contrast wherever possible.

It has not been possible to handle all variations of image characteristics in this research. Although the focus was on computerised methods of bone age assessment, the primary objective was to investigate these methods to identify the role of the carpal bones in older children. Some working assumptions were adopted in order to focus on this primary objective and to make best use of the computed radiography data available for use in this research:

- The hand can be oriented in any direction on the radiograph,
- The radiograph is of the left hand,
- All five digits are within the radiograph and all bones are visible,
- All carpal bones are within the radiograph and there are no more than eight bones,
- The radius and ulna have at least 1 cm of diaphysis visible on the radiograph,
- The radiograph is of suitable quality for reporting by a radiologist. Furthermore, it should be free of artefacts that cross the bone boundaries and free of complicating factors such as fusion of the hamate and capitate bones. These special cases will be dealt with in future research.

The demarcation between preprocessing, region of interest extraction (ROI), and the subsequent stage of bone segmentation is vague. This reflects the dependencies between these methods and the use of some segmentation methods in ROI extraction. This chapter addresses the preprocessing tasks of artefact identification and hand soft tissue outlining. The chapter begins with a review of existing preprocessing methods that have been used for automated assessment of bone age. These methods include radiographic border detection,

contrast enhancement, image artefact identification and removal, and region of interest extraction. Hand outlining has been implemented in this chapter because it is required for the identification of radiographic markers and for noise estimation methods used in Section 6.3. From the review of existing methods, region of interest extraction for automated bone age assessment is now only a development and implementation task. For this reason it has not been addressed in this research and manual methods have been used instead (please refer to Appendix D for details of the software used for these manual methods). Although this is a time consuming approach, it has the advantage of removing the influence of region of interest extraction from the carpal bone segmentation and feature extraction results (this task can be completed very reliably by a human with little training). This approach is also in anticipation of incorporating the carpal bone analysis into an existing system for automated bone age assessment.

5.2 Existing preprocessing and region of interest extraction methods

This section reviews existing methods of preprocessing and region of interest extraction that are commonly used in systems for automated assessment of bone age. Image radiographic marker identification is discussed in Section 5.3 and denoising of images is discussed in Chapter 6.

5.2.1 Detecting radiographic borders

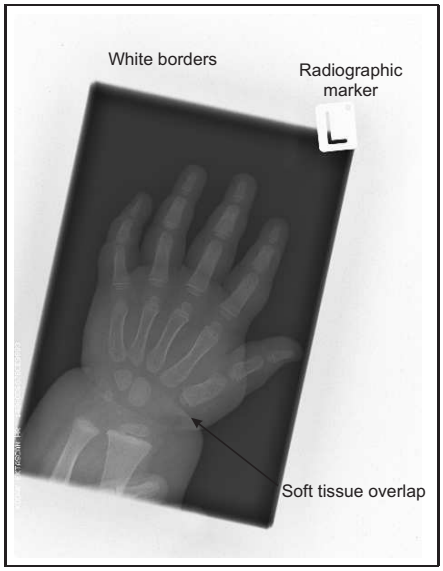
During a hand-wrist x-ray the image receptor is often exposed right to its edge. Sometimes collimators are set on the x-ray machine such that parts of the image receptor are unexposed to the primary x-rays[1]. The boundary between exposed and unexposed areas is often unsharp because of geometric and attenuation limitations of the collimators, and because of x-rays scattered from the hand. On most modern CR systems it is possible to exclude the unexposed area of the image either manually or as part of the automatic image enhancement. The unexposed area may be left as 'white' or the image changed to 'black' to reduce unwanted glare when viewing the images. Different radiographic techniques and different processing selections mean the nature of the radiographic border can be unpredictable. Figure 5.1 shows three different hand-wrist radiographs with three different radiographic borders for the same CR system. The method of finding the radiographic border needs to manage these three possibilities.

These methods have ranged from simple pixel-based searching through to more complex procedures involving such methods as the Hough transform for identifying lines in the

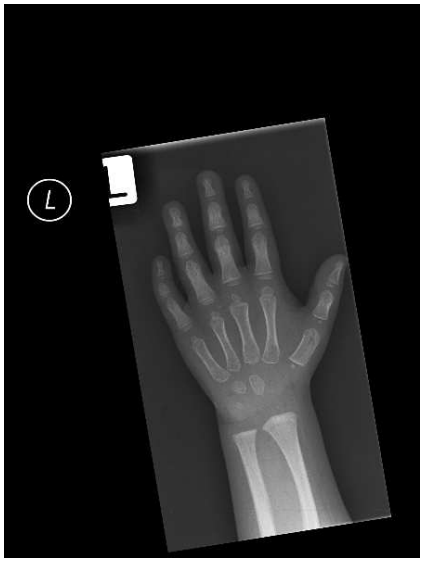
[1]X-ray beam collimation is considered good practice because it reduces unwanted scattered radiation, as well as reducing unnecessary radiation dose to the patient.

image. It is possible to search for vertical and horizontal lines that contain an empirically determined portion of pixels with intensities less than some threshold, but this only works when the radiographic borders are parallel to the edges of the image receptor [Hill94a]. Alternatively, because the standard deviation of pixel intensities in the unexposed area is less than in the main area of the image (because the unexposed area is more uniform), this can be used to identify the borders of the unexposed area. The borders can be further refined by fitting lines between pairs of potential corner points and testing pixel intensity deviations along these lines [Piet91].

The Hough transform is a method of transforming image objects into a parameter space in which it is easier to group objects to identify them. One of the simplest and most common applications of the Hough transform is to find lines in binary images. The object pixels in the binary image are converted into polar coordinate curves that are summed in a Hough space image [Prat01, p570]. Where the curves intersect in Hough space there are local maxima, and the polar coordinates of the local maxima define straight lines in the image. The largest pixel intensities in the Hough space image correspond to the dominant straight lines in the image. A refinement of this method is to calculate the gradient magnitude and direction for the image pixels, and to weigh the contribution of each image pixel to the Hough space image based on the gradient magnitude [Behi02]. By imposing further geometric constraints on the selection of points in Hough space, it is possible to identify the radiographic border on hand-wrist radiographs [Behi02].



(a) A white radiographic border and image artefacts.



(b) A dark radiographic border generated by the computed radiography system. Note the radiographic 'L' marker, as well as a 'L' marker generated by the computed radiography system.

(c) No radiographic border, but note the artefact on the right of the image caused by identification markings for the CR plate.

Figure 5.1 Different types of radiographic borders seen on computed radiography images of the hand and wrist. The computerised methods should be able to cope with all three types of borders. These images are all from a Kodak CR 900 system.

Although it appears to be a seemingly simple visual task, identifying the radiographic borders was shown by one group to require complex statistical descriptions, adaptive parameter adjustment, and estimation rules to adequately remove the unexposed background on CR images [Zhan97]. An often overlooked influencing factor is the person performing the radiograph. Poor hand positioning or poor use of collimation can significantly change the demand placed upon the method for border identification. This may explain why some researchers have only required relatively simple methods, while others have required more sophisticated and robust methods of identifying the radiographic border.

5.2.2 Background correction and soft tissue variations

One of the most common preprocessing tasks is to remove the influence of a non-uniform image background. The most common cause is the radiographic heel effect. This effect arises from attenuation of the x-rays that penetrate the anode when leaving the x-ray tube [Hend02, p81]. The resulting directional attenuation causes a non-uniform emission from the x-ray tube and therefore a non-uniform background in the image. Image segmentation methods that use analysis of the histogram of pixel intensities for binary grey-scale thresholding [Prat01] usually rely upon the objects in the image having regions of homogeneous pixel intensities. A non-uniform image background can disrupt the homogeneity of regions and will cause failure of image segmentation methods.

The most popular method of background correction is to model the background as a smooth variation across the image and to either use this model in the segmentation, or subtract the background component from the image. The challenge with the background modelling approach is that the background pixels relating to the direct exposure area have to be located to analyse them - the background region is required prospectively. This is a conundrum - to model the background it is necessary to determine the background region.

Properties of the image pixels have been used to identify background regions. One of the earliest methods was to find the direct exposure area of the image by looking at the standard deviation of pixel intensities in small regions, and to threshold the image based on the mean pixel intensities these areas [Piet91]. A similar method using criterion based on the mean and standard deviation of the pixel intensities has also been used in a recent publication [Gior07]. An alternative method is to divide the image into four quadrants and to identify the background pixels from the most frequent pixel intensity in each of the quadrants [Piet01b]. The location of the four frequent pixel intensities are then used with bilinear interpolation to produce a background image that is subtracted from the hand-wrist image [Piet01b]. This approach has produced good results, with only 1% of hand extraction processes failing due to poor background removal for the tips of the phalanges [Piet01b]. An extension of this method is to use sliding regions rather than fixed quadrant regions,

and to similarly analyse pixel properties in these regions. This method uses sliding windows on both sides of the central image axis, two at the top and two at the bottom of the image with the regions able to slide vertically or horizontally along the image [Piet01a]. The location of the regions is analysed separately for the vertical and horizontal movements and background pixel intensities based on the lowest mean pixel intensity and lowest variance. The mean pixel intensity in each region is used to generate a threshold value for the rows and columns in the image using linear interpolation. All pixels with intensities below the threshold are set to zero and ignored in subsequent image processing operations. Both this method and the four quadrant method have been used in bone age assessment systems. Differences in performance have not been investigated, but both have been used in recent publications on computerised bone age assessment [Piet05] [Piet04] [Zhan06].

Nonlinear models of the background have also been used, with the argument that such models are necessary for background removal [Lee08]. The models of the background have included simple two-dimensional third order polynomials [Lee08] [Yoo07], but more complex models based on the x-ray tube geometry have also been developed and quantitatively evaluated [Behi01] [Behi02]. The problem that remains with these methods is that it is necessary to identify the background pixels. Approaches to this problem have included: iterating between background segmentation and heel effect correction until the correction converges [Behi02]; using four regions similar to the method above and iteratively identifying background regions [Lee08]; and iteratively refining the background model by removing large pixel variations from the background region [Yoo07]. Although the first approach showed good results for a large collection of hand radiographs, it is reliant upon a sufficient region of direct exposure background being present on the radiographic image [Behi02]. This could present a problem if the hand-wrist radiographs are closely collimated to the outline of the hand, either using the x-ray machine collimators, or manually using the CR image processing functions.

In some respects, the effect of the soft tissue thickness variation, and especially the soft tissue wedge-shaped hand are more important and unpredictable than the effect of background. Theoretically, compensating for differences in attenuation caused by the soft tissue wedge-shaped hand (Figure 5.2) should be easy. However, the person-to-person variability of the soft tissue component can be high, even for children of the same bone age. This is made more difficult by differences in fat content of the tissue. Because of these problems, and the problems with background removal, such corrections have not been used in this research. Instead, segmentation methods have been used that place minimal demand upon models for the background region and soft tissue variation.

Figure 5.2 A side profile of the left hand and wrist of a nine year old boy. This photo demonstrates the wedge-shaped profile of the hand with an increasing hand thickness when moving from the distal phalanges (finger tips) to the wrist. This wedge-shaped profile causes non-uniform attenuation in the hand-wrist radiographs. The picture also demonstrates what would be considered poor radiographic technique because the hand has relaxed slightly and the knuckles have risen, resulting in changes in the radiographic angle of the phalangeal-metacarpal joints, and a slight change in angulation of the carpal bones.

5.2.3 Contrast enhancement

A common method of contrast enhancement is to scale the pixel intensities so the image histogram follows some probability density function that improves the image contrast [Prat01, p253]. This method is referred to as histogram equalisation. Different types of scaling have been used, but the most common form is designed to produce a histogram that is nearly uniform. If the image contains dark regions in which perception needs to be improved, then histogram equalisation can be of benefit because it 'stretches out' the contrast of the image. The uniform histogram equalisation method has been used as a preprocessing step for hand radiographs to standardise the dynamic range of the image and ensure the background has approximately the same pixel intensity in all radiographs [Hill94a].

There are extensions to the uniform histogram equalisation methods. Early bone segmentation research claimed that the hand radiographs had a histogram with bone, soft tissue, and background groups of pixels, and that each group had a Gaussian distribution of pixel intensities [Mich89]. This research demonstrated that it is possible to stretch the image histogram to improve the separation of these distributions. This approach has also been used with the Tsai moment-preserving method [Tsai85] to calculate thresholds to separate the image pixels into the three groups, then perform histogram equalisation separately for each of the three groups [Sun94]. A slightly different approach to histogram equalisation is to base the contrast enhancement only on edge pixels [Ko95]. This method uses a gradient filter and pixel connectivity criterion to generate a preliminary estimate of edge pixels.

The edge pixels are all that is used in the histogram equalisation. This method is thought to make the contrast enhancement less sensitive to image noise and to address the trade-off between contrast enhancement and the edge degradation that usually results from applying these methods [Ko95].

A major problem with histogram equalisation is that if the image is already of 'good-quality', then histogram equalisation may actually degrade the image [Prat01, p258]. Furthermore, with many CR systems a lot of image manipulation has already been performed to produce a suitable image for display (see, for example, [Bogu95]). Although it would be convenient to be able to standardise the pixel intensities for areas of the image such as background, the limitations of histogram equalisation mean this standardisation might be of limited benefit. Often the histogram equalisation, or other contrast enhancement method, will produce more visually appealing images. But these changes will not necessarily improve image processing. For this reason, contrast enhancement has not been used in the research reported here.

5.2.4 Region of interest extraction

The purpose of the region of interest (ROI) extraction is to generate a numerical representation of areas of the image that require separate image processing for bone age estimation. With the exception of some research on active shape models [Thod09] and neural networks (below), all computerised bone age assessment methods proposed to date require the inclusion of the image processing stages of ROI extraction and segmentation. The ROI extraction is a prerequisite to bone segmentation methods because many of the segmentation methods have unsatisfactory performance when applied to the whole hand, mostly because the methods and parameters have usually been optimised for specific regions of the hand or wrist.

The distinction between ROI extraction and segmentation can cause confusion when reviewing existing methods. Some of the methods of ROI extraction use image processing that is normally reserved for segmentation. The methods of extraction are often a succession of refinement techniques with segmentation forming one part of the process. This can blur the division between ROI extraction and subsequent segmentation stages.

The majority of research on ROI extraction has been for the epiphyseal-metaphyseal regions of interest (EMROI) of the phalanges. Analysis of the EMROI are not part of this research on carpal bone analysis, but there are some aspects of the EMROI extraction that are of interest. The early stages of EMROI extraction usually involve tasks like alignment of the radiograph and labelling of areas such as the fingers and forearm. But this is where the similarities end because the location of a developing epiphysis is much more predictable

with respect to the adjoining bones than is a carpal bone, so quite different methods are required.

Some of the ROI extraction methods can be limited by normal development of the bones. For example, methods that rely upon clear separation of the epiphysis and metaphysis for EMROI extraction often fail when these bones begin to fuse [Gior07]. This is important because the point of fusion is a milestone in skeletal maturity and it should be considered in any system for the computerised assessment of bone age. Similarly, when a child is young the carpus is surrounded by a clear region of soft tissue, but as the child matures the carpal bones overlap with the radius bones and metacarpals and the soft tissue gap disappears. The carpal region becomes more difficult to extract from the hand-wrist radiograph [Fan01]. The ideal ROI extraction method should be able to identify the region at any age of the child, with the performance of the method independent of age-specific bone features. This would allow the ROI to be used with feature extraction methods that are applicable from early ossification of the bones to full fusion or maturity.

5.2.4.1 Orientation of the hand

Finding the orientation of the hand is a standard image processing function prior to extracting the ROIs, although it is sometimes inherent in the method of ROI identification. Starting from a point of few assumptions, this orientation can be as fundamental as using recognition of the radius and ulna to determine whether the radiograph is of the left or right hand [Hill94a]. However, with the exception of one group that included both hands on the radiograph (and so had to find the left hand using thresholding and vertical profile summations) [Chan03], most researchers assume that the radiograph is that of the left hand.

Depending on the size of the hand, the hand can easily be x-rayed at any orientation on the image receptor. This will depend on the policies and education of radiographic staff at the institution in which the x-ray is performed. It has been demonstrated that correction for orientation is often required because a significant portion of hand-wrist radiographs are not taken using the conventional orientation of long axis of the hand aligned with the long axis of the image receptor [Piet01a]. Correction of the alignment may mean finding a reference axis that can be used for further analysis, or rotation of the whole radiograph to bring the long axis of the hand to the vertical.

Starting with binary thresholding using the average pixel intensity of the image (after removing image data outside of the radiographic border), the orientation of the hand can be determined using simple line search strategies that look for anatomical features like the forearm and separation of the fingers [Piet91]. More complex procedures have been used to

increase the reliability of the initial orientation [Hsie07b]. However, depending on the features chosen, these methods have limitations. For example, it is possible to x-ray the wrist slightly rotated, so the phalanges and long axis of the radius and ulna are out of alignment. Similarly, the fingers can be x-rayed at a range of angles with respect to the metacarpals. What is required is a feature that is essentially invariant of normal hand positioning, but no such feature exists for the hand. However, unless there is gross clawing of the fingers or the unusual situation of an ulna-radius deviation, the third digit or third metacarpal is the best reference feature. The third metacarpal bone (Figure 5.3) is present at birth and is usually always in alignment with the proximal third digit, the capitate, and the distal radius. For the purpose of this research, this can also assist with locating the capitate bone.

Figure 5.3 Terminology for the region of interest extraction (ROI) of the radiograph and a rectangular carpal region of interest (CROI). Because of the carpal-metacarpal bone overlap, the CROI needs to include some of the metacarpal bones. The carpal region of interest is defined by one third of the length along the third metacarpal, the middle of the first phalanx, the soft tissue boundary on the side of the hand opposing the first phalanx, one one sixth of the length of the phalanx back from a best-fit of a $\wedge$ to the gap between the internal borders of the diaphysis of the radius and ulna (or the lower border of the image, whichever is the smaller).

5.2.4.2 Knowledge-based search methods

The majority of existing methods for ROI extraction are based on knowledge-based search methods. These are heuristic search methods that use prior knowledge of the soft tissue or bone anatomy of the hand and wrist. They are usually implemented as a series of successive refinement operations, starting with identification of the fingers or radius and ulna, moving on to bone identification, then finishing with delineation of individual regions of interest. Sometimes the individual regions require further refinement like correction for radiograph misalignment so they are in a standard orientation ready for further analysis [Piet04].

A common approach to initial identification of regions is to threshold the image using either a single threshold value [Mich89] [Piet91] [Hsie07b], or adaptive thresholding method [Luis03] [Zhan06] [Zhan07]. This produces a silhouette of the hand soft tissue component, but it may include some bone outline components, especially for the distal phalanges [Piet97]. The silhouette can be cleaned-up using the morphological operations of dilation and erosion to remove unwanted regions (for example, [Effo94]).

From the silhouette image, a number of techniques have been used to locate and label the major anatomical regions of the hand. For the most part, these techniques can be grouped into line profile search methods, and methods for identifying angle changes of the traced hand contour. The profile search methods scan across the hand and run-length encode step changes between black and white values in the binary image, or some other measure of change. These step changes are related to key anatomical landmarks like the edges of the fingers. By aligning the centre of the steps, it is possible to identify, for example, the phalangeal axes [Sun94] [Piet01b] [Hill94a] [Effo94]. After smoothing the image profiles within the steps, derivative operations can be used to find bone edges for subsequent processing [Mich89].

The angle-change method works by tracing the outline of the silhouette and looking for changes in contour angle with distance along the contour. Contour angle changes, like those at the tips of the phalanges, can be used to identify anatomical landmarks [Effo94] [Mahm00] [Kwab85] [Lee08] [Yoo07]. Another use of tracing the contour is to sweep across an arc, analysing the distance from a fixed point in the wrist to the silhouette outline, and looking for distance versus angle patterns that correspond to anatomical landmarks, like valleys corresponding to the short distance to the base of the phalanges [Mahm98].

The silhouette can also be skeletonised using a thinning algorithm or medial axis transform. This process results in axes that are first approximations to the centre lines for the phalanges [Gabo97] [Luis03]. A collection of search paths are created by linking these

axes and extending them to the image boundary or soft tissue-background margin. By analysing projections perpendicular to these search paths, it is possible to identify joints and extract phalangeal EMROIs [Gabo97]. However, it is not possible to find the metacarpals using this method. They have been found instead by using a circumferential search across the hand at a fixed distance along the extended axes [Luis03].

One problem with the silhouette method is that it is reliant upon the initial thresholding or segmentation of the soft tissue and bone areas. With variable image background and soft tissue thickness variations and overlap (Figure 5.4(a)), standard segmentation methods begin to fail. An alternative is to either find ridges in the image corresponding to the long axes of the fingers [Marq01], or to find bone edges using edge detectors like the Sobel filter [Piet91] [Morr94] and the Canny edge detector [Mano94] [Hsie07b].

The carpal region of interest (CROI) can be found by scanning vertical and horizontal lines of the silhouette and looking for edges and widths that relate to anatomical features like the minimum width of the proximal end of the wrist, and the soft tissue junction between the thumb and second digit [Piet93] [Zhan07]. Because fixed horizontal and vertical scan lines are used in this method, the CROI results will depend on the performance of the orientation correction method. To some degree, this orientation limitation can be overcome by using a method that starts by finding a third metacarpal reference axis, then uses empirical CROI ratios and skin boundary regression lines to extract the CROI [Voge00].

Instead of using the silhouette to find soft tissue landmarks, it is possible to delineate a CROI by finding bone landmarks. This will only be reliable if the bones are present across all stages of maturity. One set of landmarks that has been used is the head of the ulna and the base of the thumb [Hsie07b]. These landmarks can be found by scanning the image to find the head of the ulna bone and projecting across the bone to define the base of the CROI, then scanning to find the base of thumb and projecting across the first metacarpal to define the top of the CROI [Hsie07b].

Finally, in a method referred to as the 'sticking' algorithm, a simulated needle is stuck sideways into the hand image and a histogram of pixel intensities tracked until local thresholds have been found corresponding to the bone edges [Fan01]. This method can be used to find upper and lower bounds for the CROI. Using these bounds, the rest of the carpus can be divided into bone search regions using projective pixel-intensity integration operations. This method requires good contrast between bone and soft tissue, and only works in children up to 7 years old because after this age the carpal bone separation is too small and the extraction error increases [Fan01].

5.2.4.3 A registration approach to ROI extraction

Although the hand-wrist radiograph is a visual scene that grows in complexity with the age of the child, there remain constant components in the image against which registration techniques can be used. Although there are areas of the hand that are rigid, like the majority of the metacarpal bones, hand movements such as simple rotation of the wrist or spreading or closing of the fingers mean that rigid registration methods may fail. The idea of the image containing both rigid and elastic components was motivation for a 'wire' model with articulating joints [Mart09] [Mart03]. In this model, the wires link landmarks in the radiograph that require rigid registration methods (namely, the long bone structures). Elastic registration is applied to the remaining soft tissue structures, with the elasticity proportional to the distance these structures are away from the wires.

The carpal bones are excluded from this registration because of difficulty in using the wire model with the bone sizes, and problems with finding age-invariant landmarks [Mart09]. However, the landmarks for each finger bone and metacarpal, plus the radius and ulna landmarks, are used to constraint the registration of the carpal region by joining landmarks on the metacarpals with landmarks on the radius and ulna. Unfortunately, there is significant distortion of the carpal bones because of the elastic registration.

The registration method is very detailed and complex, and even includes a correction for the influence of bone width. However, the method appears to deform the image elastically as part of the registration, with the aim of registering the image with a fixed template [Mart09]. This ties the method to the accuracy of the chosen template. It may become clear in future publications why this approach has been adopted.

5.2.4.4 The use of neural networks

In a completely different approach compared with the previous methods in this section, neural networks have been used for ROI extraction. Neural networks were used with a neural 'retina' that feed into a set of focusing networks that extracted the carpal bones [Rucc95]. In a similar method, feature extracting neural networks were arranged on a square lattice, with the lattice able to deform elastically so as to shift the focus of attention of the network processing [Bocc03]. The first approach produced a collection of centroids that were used to produce square ROI with the ROI size determined *a priori* based on the average bone size. This method was able to label the carpal bones. In the second approach, different input preprocessing and network configurations were used for the phalanges and carpal bones. Rather than perform region of interest extraction for subsequent segmentation and feature extraction, in both of these approaches the authors feed the extracted regions into a classification neural network for a bone age estimate for the each bone.

5.2.4.5 Identification and extraction using high-level processes

Most of the methods discussed so far in this section use low-level image processing functions. In some cases higher-level knowledge has been integrated into the methods as prior knowledge. There is a collection of methods that still use such higher-level knowledge, but incorporate it into techniques that provide some reasoning about the status of the ROI extraction. These methods allow multiple techniques to be applied to the same image, but as separate processes. An example is the use of an object-oriented approach that includes inter-process communication and high-level supervising processes that communicate to verify the ROI extraction, and to report any verification errors [Shim97]. In this example there was a shape model for landmark identification and matching that was defined using bone contour segments decomposed into a space that described the arclength and orientation. The use of high-level processes would be worth considering for a full bone age assessment system.

5.2.4.6 Comments on ROI extraction

The performance of ROI extraction methods have been published by only a small number of researchers. Most of these results are for the EMROI of the phalanges [Gert07b] [Piet01a] [Yoo07]. The results published for carpal region extraction suggest that the methods are reasonably reliable, with error rates of 6% for the CROI [Piet95], and 9% for carpal bone identification in young children [Fan01]. The lack of published results may be due to the performance of the ROI extraction methods being assessed intrinsically in the overall performance of segmentation or bone age assessment methods. However, without a measure of the performance of the ROI extraction, it leaves the open question of what is influencing possible poor performance: these methods, or subsequent processing steps.

There are disadvantages with both the soft tissue and bone-based methods of ROI extraction. When the contrast is low between the soft tissue and image background, the silhouette method can begin to fail and the bone-based segmentation is preferable. However, the advantage with the soft tissue method is that if the outline can be generated, the bone contours in normal hand-wrist radiographs are constrained to fall within this outline. This is important for carpal bone detection in young children because it constrains the detection region for early bone ossification.

It is clear from the review of existing methods that the task of region of interest extraction is essentially a development rather than a research one. This task is an important part of any bone age assessment system and can influence the overall automated results. However, because it is anticipated that the carpal bone analysis will be incorporated into an existing bone age assessment system, the region of interest methods were not developed as part of

this research and a manual method was used instead (see Appendix D for details of the software developed for the manual extraction). The manual method has the additional advantage of removing the influence of the region of interest extraction upon the carpal bone analysis, assuming the person performing the manual extraction can accurately perform this task – in this case it was the author who performed the task for all radiographs in the research dataset.

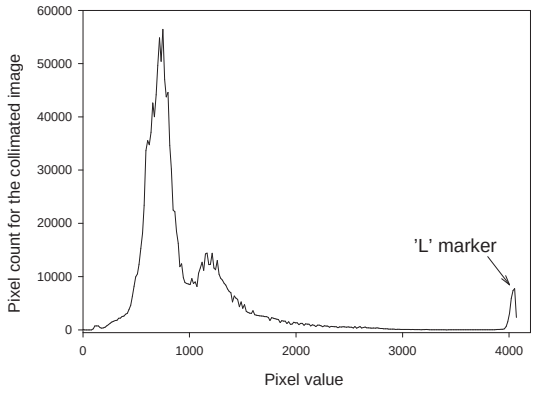
5.3 Image artefact identification and removal

Image artefacts such as lines, spots, hand restraints, patient identification markings, and position markers (Figure 5.4(a)) cause complications for image processing tasks. It is usually important to identify their location in the image and to either crop them out of the image, or to tag their location so they are excluded from image processing tasks.

With digital systems, patient identification is usually applied as an image overlay during display of the image. In this situation it does not affect image processing. The position marker is different. Although on some systems it can be applied electronically, whether it is done this way or by means of a physical marker placed on the image receptor, it usually ends up as a permanent mark on the image and is an unwanted artefact for the purpose of image processing. The positive aspect of this artefact is that the markers usually highly attenuate the x-ray beam and produce image pixel intensities at the extreme end of the pixel intensity distribution. This means that the markers are relatively easy to identify, as illustrated by the pixel intensity histogram in Figure 5.4(b).

The approach to removing small spot and line artefacts has mostly been the use of knowledge-based morphological operations and elimination of small regions through prior knowledge of typical bone sizes. This elimination is usually performed as post-processing of the segmented images and may be combined with other methods of bone object refinement [Zhan06].

(a) Artefact image for a five month old child. This image shows the 'L' radiographic marker, and a compression pad across the fingers and wrist - to hold the child's hand still.



(b) Pixel intensity histogram for the radiograph, clearly showing the pixel intensities for the 'L' marker. The histogram shows the frequency distribution of pixel intensities. This histogram clearly shows the separation between the pixel intensities for the radiographic marker and those for the rest of the image.

Figure 5.4 A hand-wrist radiographic image, with a number of artefacts, and its corresponding pixel intensity histogram.

The other characteristic of the radiographic marker that helps with the detection is that the markers are usually uniform areas on the image. From the results presented in Figure 4.11 it is clear that high intensity pixels have high levels of noise. If the radiographic marker is artificially generated by the CR processing system, then it will have no noise, although it can have lower intensity pixels around its border depending on the compression employed for the image storage. If it is generated by the use of a radiographically opaque marker, then there will be radiographic noise, and the level of noise can be relatively high. This will be indistinguishable from high intensity pixels associated with areas of high bone attenuation. However, it is possible to distinguish between the two using an initial estimation of the outline of the hand and only classifying a region as belonging to a radiographic marker if it is outside of the hand soft tissue area. The situation of the marker being inside the hand soft tissue outline has not been addressed in this research.

In this section the radiographic position marker artefact is investigated. The handling of the remaining artefacts has been reserved for future research and radiographs with such artefacts have either not been included in the research image dataset, or have been observed to have an insignificant effect upon the image processing. Impulse noise identification and removal is investigated in Section 6.2 and this is applied before any further methods of artefact identification or removal are applied.

5.3.1 Using the pixel intensity histogram

The first thing to note from the hand-wrist radiographs is that there is no clear demarkation in pixel intensities between the pixels associated with the bones, the soft tissue and the background. Furthermore, as illustrated in Figure 5.5, the pixel intensity histogram for the soft tissue is effectively bi-modal. There is no simple Gaussian distribution or mixture model that could be used for, say, the segmentation of these groups of pixels by single- or multi-level thresholding.

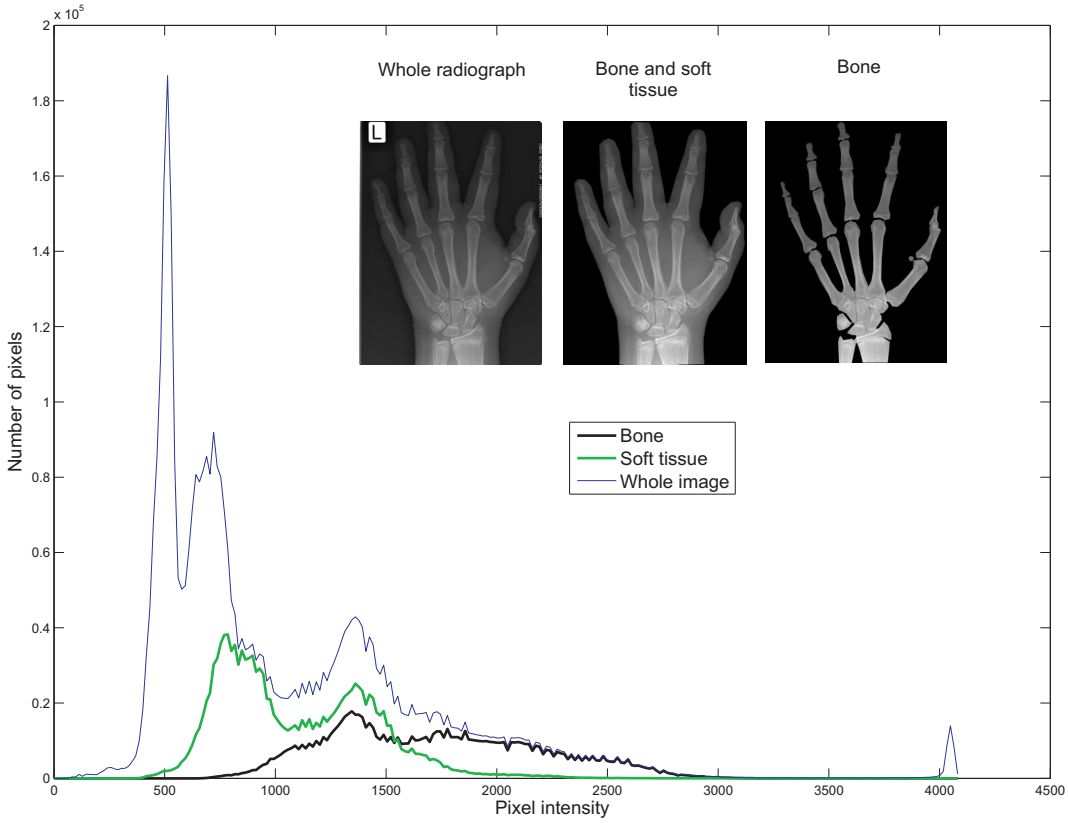


Figure 5.5 Pixel intensity histograms for the separate components of a typical hand-wrist radiograph. The histogram has been rebinned into 256 bins to reduce the noise due to pixel intensities that are not present in to CR output and are always set to zero. Pixels with intensity of less than 3 have not been included in the histogram because the CR system performs masking and usually sets the pixels outside of the mask to zero. There is no clear separation of the peaks so it would be difficult to apply, say, a Gaussian mixture model to this histogram.

To get an idea of whether the whole-image histogram results shown in Figure 5.5 are typical of clinical images, 97 radiographs were processed and the statistics were collected for each image for pixels within the collimated radiation field on the image. The mean and standard deviation across the collection of 97 images is shown in Figure 5.6. As mentioned earlier

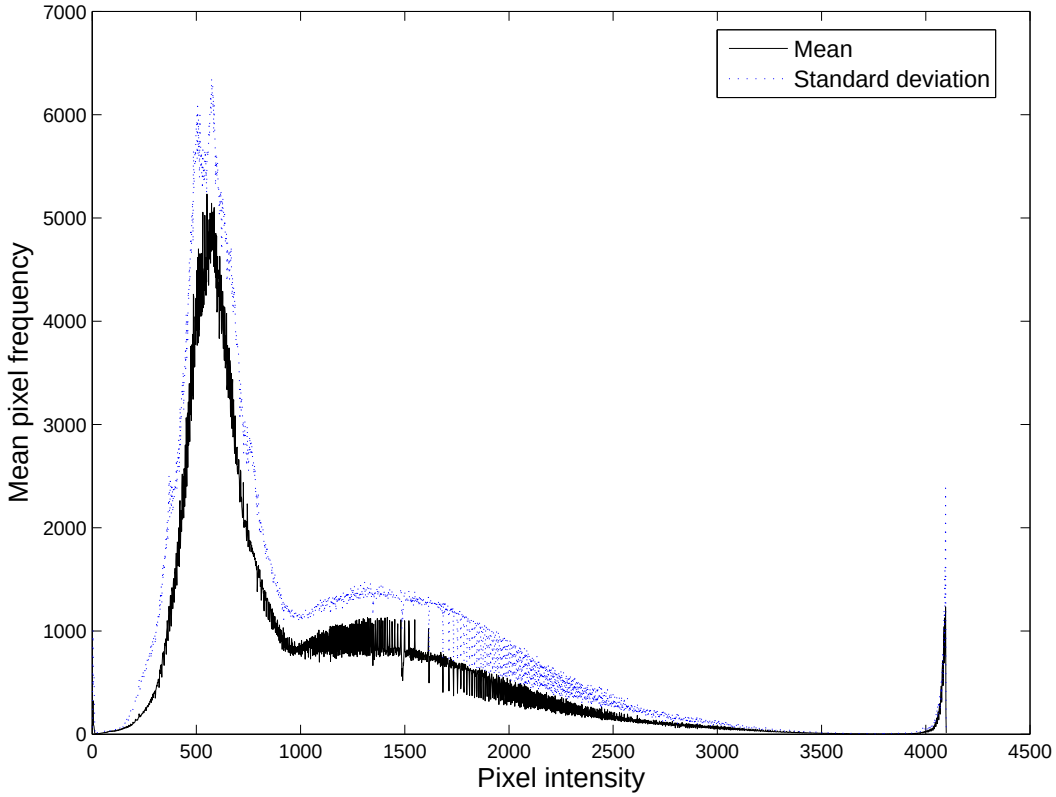


Figure 5.6 Mean and standard deviation of the pixel intensity histograms for 97 hand-wrist radiographs. Although the mean generally agrees with the results presented in Figure 5.5, the standard deviation is large so the result must be interpreted with caution. What is clear from both curves is that the cut-off of 3700 for the radiographic content is an appropriate value to use. The pixel intensities above this correspond to the radiographic markers. Note: the histogram was smoothed with a 1D Gaussian filter to reduce the noise for the purposes of presentation.

in the section, there are some parts of this histogram that are consistent and can be of use in preprocessing. Firstly, the radiographic markers tend to have high pixel intensities and these are relatively easy to remove by rejecting those pixels with intensities above 3700. There are two possible approaches to the rejection of pixels: (a) mask the pixels and perform inpainting to smooth away the effect of the missing pixels, or (b) tag the pixel with a special number so they can be identified in subsequent processing tasks and excluded from the analysis, if appropriate. The later was the approach adopted in this research with pixel values set to zero because the radiographic images never normally contain zero pixel intensities in the main radiation field. A lower limit was also set, so pixels with intensities below 100 are not of clinical significance for bone age assessment and can be excluded from any analysis.

The success of identification of the radiographic marker stems from the finding in Figure 4.5 that the CR system maintains pixel intensities across a range of x-ray doses. From

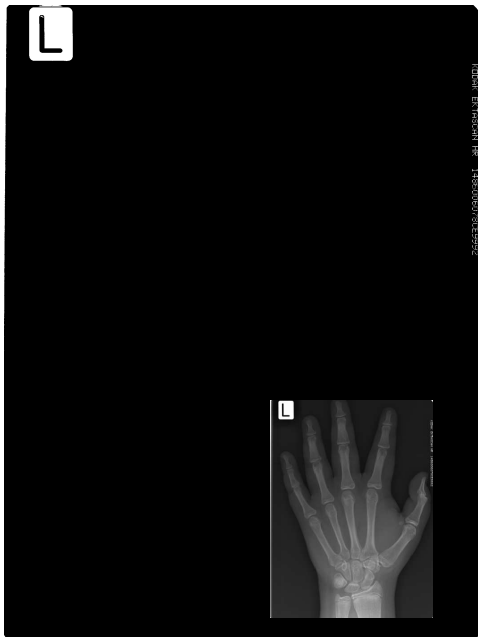


Figure 5.7 A binary mask created using the thresholds of pixel intensity < 100 and > 3700. The inset shows the original image.

this, one can be assured that the intensity of the radiographic marker will remain approximately constant across the range of clinical radiographic exposure. The exception to this is at very low dose exposures where the pixel intensities change, but as will be discussed in Section 6.3.1 these can be easily identified from the value of Exposure Index stored in the image DICOM files.

Figure 5.7 shows an example of the resulting binary mask for a clinical image where the markers and background were found using the simple thresholds of 100 and 3700.

5.4 Conclusions

From the literature review, there is a significant amount of research that has been undertaken relating to the preprocessing tasks for bone age assessment. It appears that much of this research can now be translated to development work. For this reason it was not investigated as part of this research. Instead, where necessary, manual intervention was used, but not to the point that this intervention exceeded the capability of existing methods – otherwise additional research would be required. This was so that the methods developed can be implemented in an existing system, or on a local system that is being developed.

Chapter 6

Denoising the hand-wrist radiograph

Image denoising is often considered a preprocessing task, but in this research it has been considered separately because the results of the preprocessing are used for noise estimation as part of the denoising algorithm. This chapter introduces a single-image method of estimating the radiographic noise in hand-wrist radiographs. This estimate is used to set parameters for image denoising. The denoising method has been treated as a separate task from segmentation. This is in contrast with techniques that perform these tasks simultaneously [Tsai01].

6.1 Introduction and existing methods

Like most of the preprocessing methods, the requirements for denoising depend on the subsequent image processing tasks. The most important of these tasks is image segmentation. This dependency usually involves a trade-off between the denoising method and both the detectability of objects and accuracy of edge or region detection. This dependency creates problems when investigating denoising methods. Optimising the denoising method usually involves applying the method then applying the segmentation method and determining the influence of the denoising on the segmentation results. Unfortunately, this means that the results can be significantly influenced by the segmentation method, so there is no truly independent assessment of the denoising method. One of the aims of this research was to decouple this dependency by finding an alternative method of assessing the quality of the denoising, without the need for segmentation. In this way the segmentation method can be optimised for a fixed denoising method.

There is a vast amount of published literature on the topic of image denoising, with a multitude of methods investigated for many different applications. It is clear that there is currently no universal image denoising method. Many of the denoising methods are appli-

cation specific and depend on the type and significance of the noise in the original image, as well as the tolerable noise and blur of the processed image. For example, because the human visual system can cope with very low signal to noise ratios, processing an image that originally contained a high level of noise may result in an image that is visually acceptable to a human (for the task at hand), but can contain too much noise for subsequent image processing tasks. The structural content of an image also determines requirements of the denoising method. For example, periodic structures such as texture in images often requires different denoising methods than regions of relative homogeneity. Similarly, many natural scenes contain sharp object boundaries whereas edges in radiographic images are less sharp; the requirement for edge-preservation is different between the two.

A review of denoising methods by Buades *et al.* provides a summary and classification of recent developments in image denoising, with the exception of hybrid methods [Buad05]. Investigation of a range of these denoising methods was beyond the scope of this research, although some recent methods such as non-local means may be worth investigating in the future. The approach taken instead was to expand upon denoising methods that have already been applied to hand-wrist radiographs.

6.1.1 Existing methods of denoising hand-wrist radiographs

The hand-wrist radiographs invariably contain noise from image receptor stochastic processes, image receptor structural noise, electronic noise and, although technically not a noise source, an apparent noise from the radiographic projection of the internal composition of the bones: the bony trabeculae. These noise sources can decrease the homogeneity of objects and have a significant impact upon image processing operations such as edge detection. The importance of denoising for successful segmentation of screen-film hand-wrist radiographs was recognised some time ago [Mano94]. In many respects this has not changed with the introduction of digital x-ray systems. In digital systems such as computed radiography or direct radiography there is often little control over the intensity of the final image because of automatic image manipulation. This means that a low x-ray exposure is typically adjusted to provide an acceptable image contrast and, with the exception of an increase in noise in the radiograph, there are few clues to a low x-ray exposure[1]. Unfortunately, a low x-ray exposure results in increased image noise and a reduction in the detectability of small contrast differences. With screen-film radiographs a low x-ray exposure leads to a light, non-diagnostic radiograph that has to be repeated. With the automatic image adjustment of digital x-ray systems, and the remarkable ability of the human visual system to cope with image noise, it is tempting to not reject low-exposure, high-noise hand-wrist radiographs. This means that denoising methods are at least as important for digital

[1] Most digital x-ray systems do offer a numerical measure or visual indicator of the acceptability of the x-ray exposure, but once submitted to the PACS, this measure or indicator is usually not immediately obvious.

systems as they are for screen-film radiography. Many CR systems include a facility for image denoising in the clinical processing protocols, but if these are used they are usually set to provide a visually acceptable image rather than one that is suitable for image processing.

One of simplest denoising methods that has been used with the hand-wrist radiographs is the mean filter [Hill94a]. This filter is a neighbourhood operation that replaces the central pixel of a sampling neighbourhood with the mean intensity of the pixels in the neighbourhood. This produces a linear low-pass filter that attenuates the high spatial frequency components of the image; the components that often arise from the lack of spatial correlation of image noise [Prat01, p262]. Although the low-pass filter operation works well for some noise sources (especially additive uniform noise and Gaussian distributed noise [Prat01, p269]), its largest drawback is the blurring of important image features like edges. To get around this limitation, and recognising that some of the noise components in hand-wrist radiographs are more impulse-like, some researchers have used nonlinear filters. The most popular of these is the median filter, a similar operation to the mean filter except the central pixel is replaced with the median of the neighbourhood pixel intensities [Prat01, p271]. This impulse suppression filter has been used for additive noise removal both in isolation [Dine95], and in combination with a Lee filter [Piet05] [Piet04] and a refined Lee filter [Gert07a]. The Lee filter uses local image statistics and a mean-square error estimator to produce a noise reduction method [Lee80]. The refined Lee filter takes this a step further and adaptively constrains the local statistics calculations in neighbourhoods that have high image gradients (edges) [Lee81]. It is claimed that this approach reduces the image noise, but with less blurring of edges than with the standard Lee filter [Lee81]. However, it has been demonstrated that the Lee filter is not optimal for signal-dependent noise [Kuan85], as is the case for radiographic images.

The refined Lee filter improves the preservation of edges by using the gradient or edge direction in calculations for the degree of local noise suppression. This is similar in principle to the method of anisotropic diffusion that has been used for the denoising of hand-wrist radiographs [Zhan07] [Lee08] [Lin05]. Anisotropic diffusion filtering (see Section 6.4) is a method in which the degree of smoothing (diffusion) at a point in the image is determined by a function of some image property at that point. The usual image property is gradient and the function is chosen so that the image smoothing is smallest in directions with high gradients, and largest in directions with low gradients. On hand-wrist radiographs this means that bones edges are smoothed the least (edges are preserved) and large areas with low contrast, such as soft tissue areas, are smoothed the most. Although anisotropic diffusion achieves edge preservation that is beneficial when using edge-based segmentation methods [Zhan07], in its standard form it can suffer from 'bleeding' of the smoothing operation through low contrast edges, producing a smooth hole in the bone contour.

Gabor filters have also been used for denoising of hand-wrist radiographs. The Gabor filters are a collection of two-dimensional sinusoidal filters that have a Gaussian-shaped envelope [Prat01, p544]. The orientation and frequency of the sinusoidal variations are controlled by the filter parameters and this means that it is possible to change the degree of smoothing in specific directions. For example, one group of researchers used four different filter orientations, resulting in four separate filtered images. These filtered images were then segmented separately before being combined to produce a single segmented image [Chan03].

The carpal region of the hand-wrist radiograph often contains bones with a lot of bone structure that appears on the radiographs as a type of textural noise. It is termed noise here because it is too inconsistent between children to be of use as a texture for segmentation. Although it is necessary to as far a possible preserve the edges of the bones during denoising, it is not necessary to preserve the edges of this bone structure. For this reason, only the bone contours have been considered for the evaluation of denoising algorithms.

The approach chosen for denoising begins with the removal of impulse noise using a robust yet simple local statistic of pixel intensity variability. This is in contrast with some more elaborate methods of impulse noise detection and removal. The noise is then estimated using local statistics and the exponential noise model developed in Chapter 4. The signal-dependent noise model is then used in a refined anisotropic diffusion method to denoise the hand-wrist radiographs.

6.2 Impulse noise reduction using median filtering and outlier detection

The first artefact to remove from the images is impulse noise because of the impact it can have on subsequent image processing task. The most common cause of impulse noise is dust or dirt on the computed radiography (CR) plate. The dust or dirt reduces the stimulated emission of light from the plate and produces a white spot on the radiograph. These impulse noise spots are very localised and have a high contrast. From inspection of a collection of 2392 x 1792 pixel Kodak CR images, the white spots usually involve from one to ten adjacent pixels, with the most common being three.

Linear filtering is a fundamental method in signal and image processing, and although it is good at suppressing Gaussian-type noise, it does not work well for impulse-type noise [Jahn02, p307] [Riou96]. There are a number of methods of impulse noise reduction that can be applied to radiographs [Niko04] [Wind01] [Garn05] [Fros09], but the most common method is to use a neighbourhood filter such as the star-shaped (4-connected)

median filter [Piet01a] [Piet04] [Gert07a]. The median filtering is usually followed by filtering to reduce Gaussian noise, but the impulse noise removal and Gaussian denoising can also be combined into one method [Garn05]. This combined method has not been used in this research because of the complexity of optimisation of the method. In this section the star-shaped median filter is compared with two outlier-detection filters: one based on a sample standard deviation measurement, and the other using the median absolute deviation (MAD) robust estimator of statistical dispersion. The performance of the three methods was tested for impulse noise removal using the image quality metric of mean Structural Similarity Index (SSIM), an alternative to mean-squared error (MSE) [Wang04].[2]

6.2.1 Median and MAD filtering for impulse noise removal

The problem with linear filtering is the inability of the method to adapt to image content. Many boundaries of objects in an image act like discontinuous functions and it is these features that are often of interest. It is important that denoising methods be able to distinguish these discontinuities from noisy oscillations and this requires nonlinear filtering [Chan05, p70]. The starting point of nonlinear filtering is median filtering, an example from the class of order-statistics filtering.

Arias-Castro and Donoho present a mathematical treatise of median filtering with some interesting discussion on the performance of linear filtering and median filtering [Aria09]. They challenge what they call the median folk theorem that "Median filtering outperforms linear filtering for suppressing noise in images with edges". Their conclusion was that based on a decision-theoretic framework, the median folk theorem is false except in the case of negligible noise level per pixel. Interestingly, for a low frequency of corruption of an image by impulse noise, the mean noise level per pixel is very small, although it is unclear whether this finding relating the median and linear filtering applies to impulse noise.

Let $N_j(x)$ be a $(2n+1) \times (2n+1)$ neighbourhood of pixels centred at the pixel x in an image, and let I_j be the corresponding pixel intensities of this neighbourhood (please refer to Figure 6.1) with the intensity of the central pixel $I(x)$, then the median filter sweeps this neighbourhood across the image and replaces the central pixel with the intensity

$$I(x) \leftarrow \underset{j \in N(x)}{\text{median}}(I_j)$$

[2] MATLAB® code for the SSIM index was downloaded from www.ece.uwaterloo.ca/~z70wang/research/ssim/ (accessed Jan 2010) .

For $s = (2n+1) \times (2n+1)$ and $I_j \in \{I_1 \leq I_2 \leq \cdots \leq I_s\}$

$$\underset{j \in N(x)}{\text{median}}(I_j) = I_{\frac{s+1}{2}} \quad \text{because } s \text{ is odd.}$$

The median operation is nonlinear because for a distribution k scaled by factor s,

$$\text{median}(s \cdot k) = s \cdot \text{median}(k),$$

but for two distributions k and m,

$$\text{median}(k+m) \neq \text{median}(k) + \text{median}(m).$$

This nonlinearity means that the median filter cannot be implemented as a convolution operation.

The star-shaped median filter uses a more localised 4-connected pixel neighbourhood with a '+' shape that includes the central pixel and the four adjacent pixels [Piet04]. The median calculation for the new central pixel intensity $I^{new}(i,j)$ is

$$I^{new}(i,j) = \text{median}\{I(i+1,j), I(i-1,j), I(i,j+1), I(i,j-1)\}.$$

For a Gaussian distribution the median is equal to the mean. The mean is not a robust estimator, whereas the median is a robust estimator of central tendency in that it will tolerate outliers such as impulse noise. If there is an impulse in a sample neighbourhood then the mean will be affected by this value. The median of the sample will not be affected until 50% of the sample values are outliers. This 50% is referred to as the breakdown value for the robust statistic: the number of outliers the sample can contain before the estimate becomes arbitrarily large [Rous02]. This outlier tolerance means that the central pixel is replaced with a robust estimate of the mean of the pixel intensities in the neighbourhood, and, hence, is less influenced by extreme pixel intensity changes in the neighbourhood, such as happens near edges of objects. This leads to the ability of the median filter to preserve edges compared with, say, a mean filter.

The drawback with the median filter is that it replaces every central pixel of a neighbourhood with the median of the neighbourhood. This causes noise suppression, but can also lead to signal suppression [Prat01, p271]. Furthermore, this changes the noise characteristics of the image and the requirement in this research was to remove the impulse noise with minimal impact on the overall image noise so that the exponential noise model of Eqn. (4.6) can be used for noise estimation. The approach used in this research was to separate the tasks of detecting and removing impulse noise, effectively using a 'switching

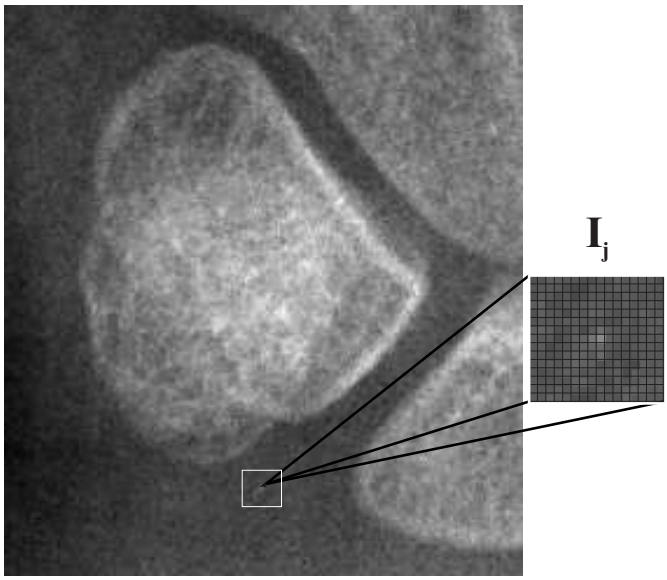


Figure 6.1 The pisiform bone with a square neighbourhood for filtering. The neighbourhood window sweeps across the whole image, replacing the central pixel with a value based on some calculation using the neighbourhood pixel intensities I_j. The neighbourhood is defined by its size, shape, and the location of the central pixel.

filter' approach [Riou96] [Fros09].

The outliers are detected using a statistical discriminator based on finding the sample standard deviation s of a deleted neighbourhood and calculating the absolute difference between the central pixel intensity $I(x)$ and the mean of the deleted neighbourhood[3]. If this difference exceeds a factor t times the standard deviation, then the central pixel is classified as an outlier and is replaced by the median of the deleted neighbourhood. Although the impulse detection was only meant to be applied to positive-going impulse noise, an absolute value operation was used to detect and remove negative-going impulses as well. This had the effect of removing statistical outlier pixels with intensities less than typical impulse noise. This effectively tidied-up the local statistics of the neighbourhoods by correcting for excessive pixel fluctuations. For $t = 3$ and a Gaussian distribution of pixel intensities, only 0.26% of the central pixels will be replaced by the median value of the deleted neighbourhood if no impulse noise is present. The algorithm is listed in Algorithm 1 on page 118.

This method works well if only the central pixel of the neighbourhood has been corrupted by impulse noise. If other pixels have been corrupted then the standard deviation increases significantly for the deleted neighbourhood because of the quadratic term in the standard deviation calculation $s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2}$ where x_i are the samples and N is the

[3] A deleted or punctured neighbourhood in this case is the set of pixels in the neighbourhood omitting the central pixel.

number of samples. The median absolute deviation about the median (MAD) given by

$$\mathrm{MAD}(y_i) = \underset{i}{\mathrm{median}}\{\Big|y_i - \underset{j}{\mathrm{median}}(y_j)\Big|\} \tag{6.1}$$

is a robust estimator that is an alternative to the standard deviation for measuring statistical dispersion [Rous87]. It is very robust with a breakdown value of 50%. The standard deviation estimate $\hat{\sigma}$ is calculated using $\hat{\sigma} = \frac{1}{k_n} \cdot \mathrm{MAD}$ where k_n is a scale factor that depends on the distribution and the number of samples [Foi08]. For a Gaussian distribution with $n > 20$,

$$\hat{\sigma} = \frac{1}{k_n} \cdot \mathrm{MAD} \quad \text{and} \quad k_n \approx \frac{1}{5n} + \Phi^{-1}(\tfrac{3}{4}), \tag{6.2}$$

where $\Phi()$ is the inverse of the cumulative distribution function for the Gaussian distribution [Foi08].

By replacing the mean with the median and the standard deviation with the MAD-based $\hat{\sigma}$ estimator, it is possible to implement an outlier detection method that is more robust to more than one pixel in the neighbourhood being corrupted by impulse noise. The neighbourhood operation for implementing this filter is listed in Algorithm 1. This operation is applied to all pixels in the image, with zero padding of the edge pixels of the image. Because of the robust nature of the MAD estimator there is typically little difference in practice between using a full neighbourhood that includes the central pixel, and using a deleted neighbourhood. A deleted neighbourhood was used in this research. With the MAD estimator having a breakdown value of 50% [Rous93] this means that the MAD estimator was very robust. The MAD estimator does, however, have some drawbacks and these are discussed in Section 6.3.

Algorithm 1 Neighbourhood operation for impulse noise removal using the MAD robust estimator of local statistical dispersion. This operation is applied to each pixel neighbourhood in the image.

Input: I set of neighbourhood pixel intensities, t threshold in units of sigma (typically 3)
Output: y pixel intensity to replace the central pixel

$x = I_{\textit{central pixel}}$
$I_j = I \backslash x$ // Deleted neighbourhood
$I_{med} = \mathrm{median}(I_j)$
$MAD = \mathrm{median}(\Big|I_j - I_{med}\Big|)$
// Correct MAD for small sample bias and check for outliers using threshold
if $|x - I_{med}| > t * \frac{1}{k_n} * MAD$ **then**
$y = I_{med}$ // Replace the central pixel with the median of the neighbourhood
else
$y = x$ // Return the current central pixel value
end if
return y

6.2.2 Comparison of the outlier filters

The star-shaped median filter, standard deviation-based outlier filter, and MAD-based outlier filter were tested using both samples of observed impulse noise and a clinical radiograph of the carpus. The same neighbourhood size was used for both the detection and the median calculation for removal of the impulse.

A 5x5 pixel neighbourhood was empirically found to give the best performance for outlier filtering of the hand-wrist radiographs. This choice of neighbourhood size involved a trade-off between false detection and performance of the true detection-correction. Increasing the size of the neighbourhood decreased the number of false pixel changes, but also decreased the accuracy of local pixel intensity calculations. The outlier filter with a larger neighbourhood size was less effective at impulse removal at edges, probably because of incorrect estimates of the pixel intensity dispersion for neighbourhoods with high contrast at edges. Increasing the neighbourhood size also lead to under-correction of impulses if two or more impulses were close together with respect to the neighbourhood size. If the neighbourhood size was decreased, large size impulse noise was not properly removed. A reduced neighbourhood size also resulted in an increased percentage of false impulse detections and this had the undesirable effect of changing the underlying distribution of pixel intensities.

The threshold t for the outlier filter was set to three. Increasing this value reduced the impulse detection. For bones with relatively sharp edges there was some increase in the number of holes near the edges, but there remained an overall improvement in the level of impulse noise suppression, as assessed visually.

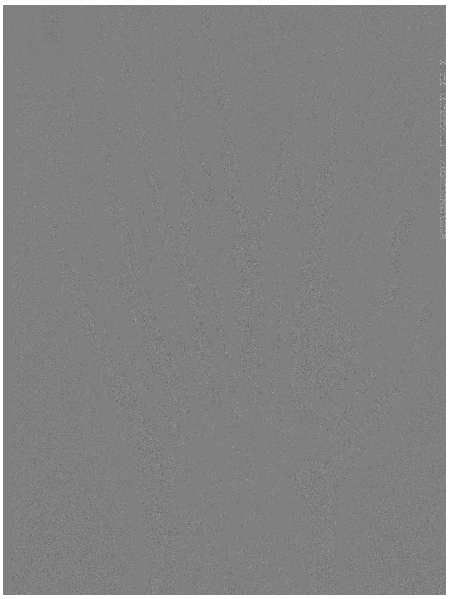
6.2.2.1 Method noise of the MAD-based outlier filter

The method noise is the difference between the original image and the denoised image (please refer to Section 6.5). Figure 6.2 illustrates the method noise for a typical radiograph of the whole hand. Important to note from the method noise image is the lack of bone edges or the bone texture that results from, say, the trabeculae bone. Figure 6.2(b) shows that both positive and negative impulse noise was removed by the outlier filter. The changed pixels are shown in Figure 6.2(c) where the changes are shown in a binary representation with white being a changed pixel. Figure 6.2(c) suggests that there have been an extensive number of pixels changed by the outlier filter, but this is a characteristic of the image display because in this image only 3.9% of the pixel intensities were changed. Even though this number of pixel changes is smaller than may be visually apparent, they are of concern because for a Gaussian distribution of pixel intensities one would expect that only 0.26% of pixels without impulse noise would be falsely detected as outliers by a outlier

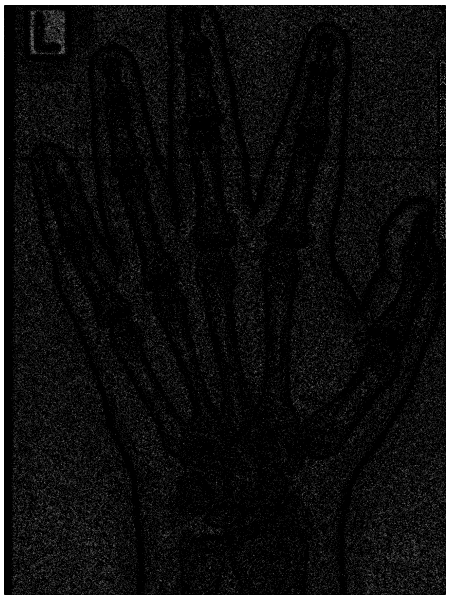
filter with $t = 3$. As mentioned in the previous section, the false detection of impulse noise was also influenced by the neighbourhood size. For example, for the radiograph shown in Figure 6.2(a) the percentage of pixels changed by the MAD-based outlier filter was 8.2%, 3.9%, 2.8%, and 2.2% for neighbourhood sizes of 3x3, 5x5, 7x7, and 9x9 pixels, respectively. Some of this can be explained by the bone edges and bone texture in the image causing false detections, but this is unlikely to be significantly influenced by neighbourhood size and is not supported by the results presented in Figure 6.2(c). A more likely explanation is that the distribution of pixel intensities in regions of the Kodak CR 900 hand-wrist radiographs are not Gaussian. This is discussed in more detail in Section 6.3.2.1, but what is important to note from Figure 6.2(c) is that the there are few pixel changes in regions of the image that contain bone edges, bone texture, or soft tissue edges. Furthermore, Figure 6.2(c) shows the binary representation of the pixel changes, whereas the overall impact on the image is best represented by Figure 6.2(b) where the false pixel intensity changes are relatively small and have little impact on the image visually.

The operation of the outlier noise removal on specific impulses is demonstrated in Figure 6.3 for a single pixel impulse, and in Figure 6.4 for an impulse noise involving five pixels. In both cases the before and after images illustrate the success of the impulse noise removal and the method noise image demonstrates the minimal impact of the filter on the non-impulse pixels.

(a) Original image



(b) Method noise for the MAD filter



(c) Binary version of the method noise

Figure 6.2 The impact of the MAD-based outlier filter on the image pixel intensities and noise. The original full resolution image, (a), was filtered with a 5x5 neighbourhood and a threshold of three. The method noise, (b), is the difference between the original and filtered image. The display range for the method noise was set at plus and minus one standard deviation of the method noise intensities. To further demonstrate the pixel changes, a binary version of the method noise, (c), was generated by highlighting the non-zero pixel intensities of the method noise. The method noise is pixel-intensity dependent, but does not have a structure caused by edges. The lack of pixel changes at the edges does, however, reflect the limitation of this outlier filtering method where there are large image gradients. The binary version of the image noise appears to contain an extensive number of changed pixels, but this is appearance is caused by downsampling of the image because only 3.95% of the pixels have been changed by the filter.

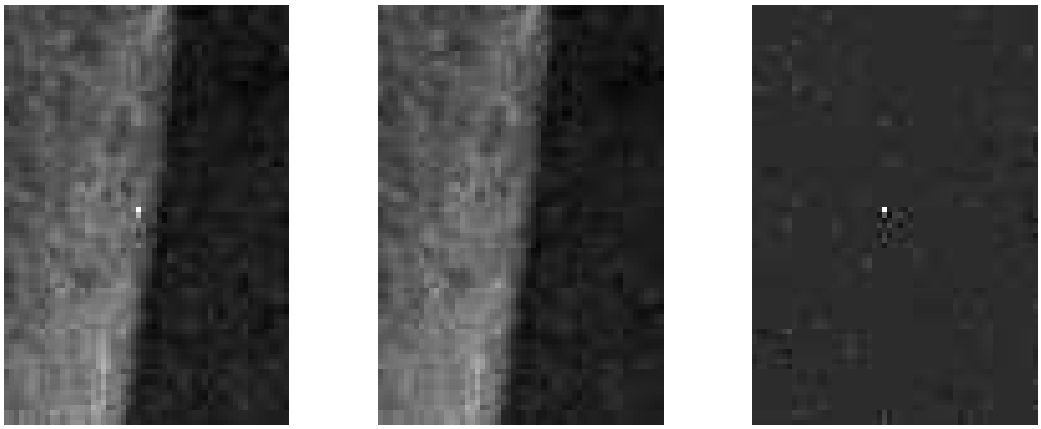


Figure 6.3 Single-pixel impulse noise removed using the MAD-based outlier filter. (a) original radiograph; (b) denoised image; (c) method noise. Note the minimal impact of the outlier detector on the image noise.

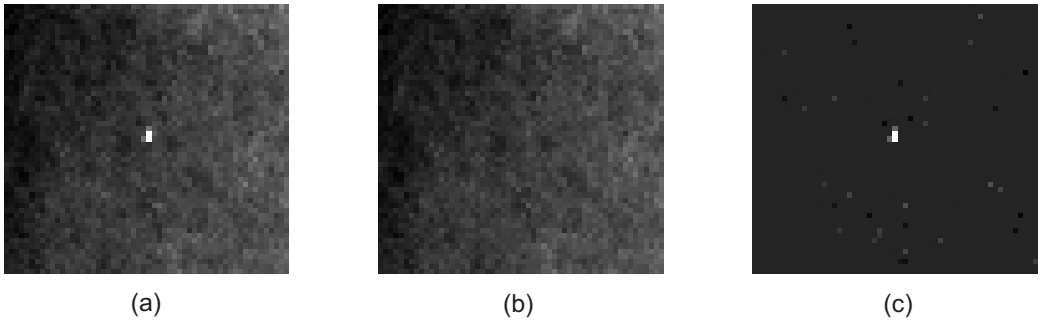
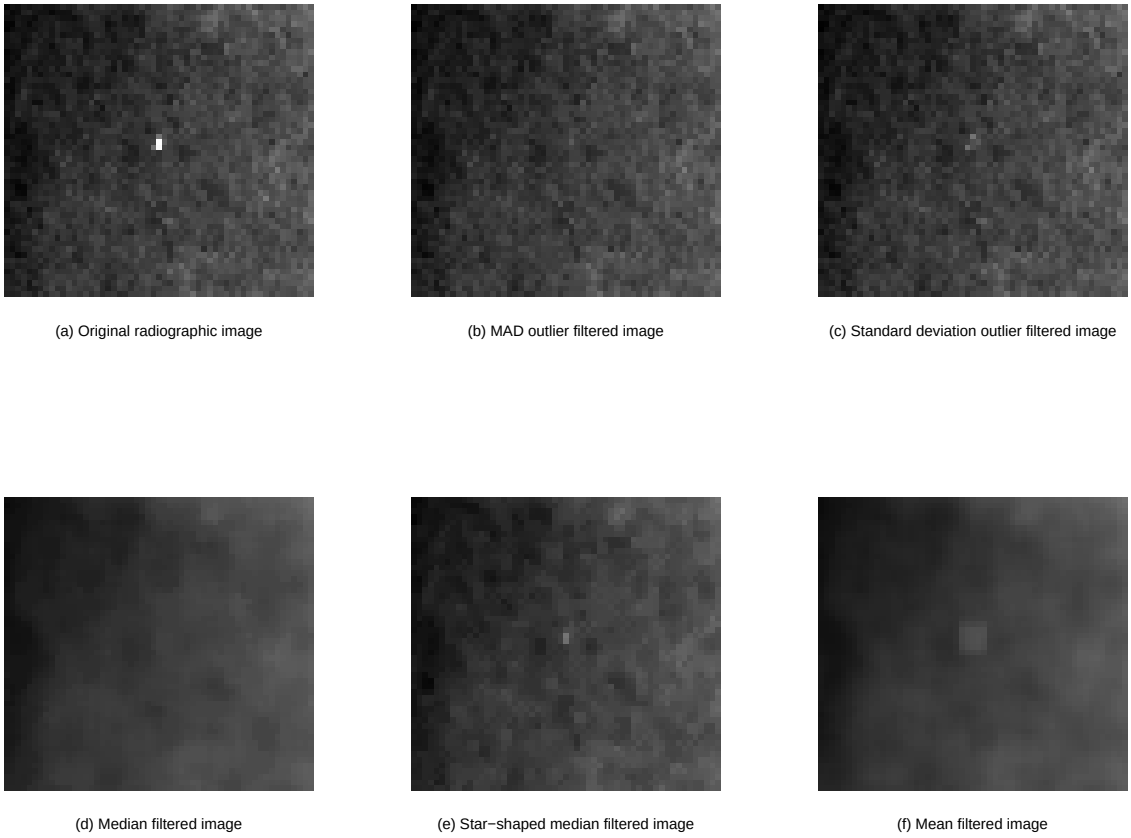


Figure 6.4 Multiple-pixel impulse noise removal using the MAD-based outlier filter. (a) original radiograph with 5 pixels corrupted by impulse noise (b) denoised image; (c) method noise.

6.2.2.2 Comparison of the different impulse noise removal methods

A comparison of the different impulse noise removal methods is illustrated in Figure 6.5 and Figure 6.6 for two clinical images. Both the MAD-based outlier and median filters remove the impulse noise from the original radiograph of Figure 6.5, but only the MAD-based filter leaves the image noise essentially untouched. Both the median filter and mean filter significantly alter the image noise and cause image blur, as illustrated in Figure 6.6(d) and (f). The significant residual impulse noise in the epiphyseal gap of the radius bone for the standard deviation based outlier filter (Figure 6.6 (d)) most likely reflects lack of robustness of the standard deviation calculation compared with the MAD estimate of statistical dispersion. The standard deviation calculation for a neighbourhood containing only a few large image gradients will be larger than with a MAD-based estimate of dispersion, mean-



(a) Original radiographic image

(b) MAD outlier filtered image

(c) Standard deviation outlier filtered image

(d) Median filtered image

(e) Star−shaped median filtered image

(f) Mean filtered image

Figure 6.5 Comparison of the different outlier filters. The MAD-based outlier filter (b) performs well at removing the impulse noise, as does the standard-deviation based filter (c), although there is some residual impulse noise. The median filter removes the impulse noise, but significantly changes the noise characteristics. The star-shaped median filter (e) has less impact upon image noise, but does not completely eliminate the impulse noise. The mean filter (d) alters the image noise and does not completely remove the impulse noise.

ing that it is unlikely the difference between the central pixel and the neighbourhood will be calculated as statistically significant, so the impulse noise will not be corrected. The limit of performance of the MAD-based outlier filter is illustrated in Figure 6.6 (b) where impulse noise on the large edge gradient of the ulna bone (left-hand bone in the image) is not completely eliminated by the filter.

The outlier filter methods were assessed quantitatively by adding impulse noise of different shapes to a clinical radiograph of the carpus that had the MAD-based and standard deviation-based outlier filtered already applied to it in separate experiments. The amount of impulse noise removal was measured using the mean Structural Similarity Index (MSSIM) [Wang04]. This index provides a measure of difference between a test image and a reference image and is an alternative to the mean square error. The index is calculated as a mean value $\mathrm{MSSIM} \in [0, 1]$. The index combines a comparison of luminance,

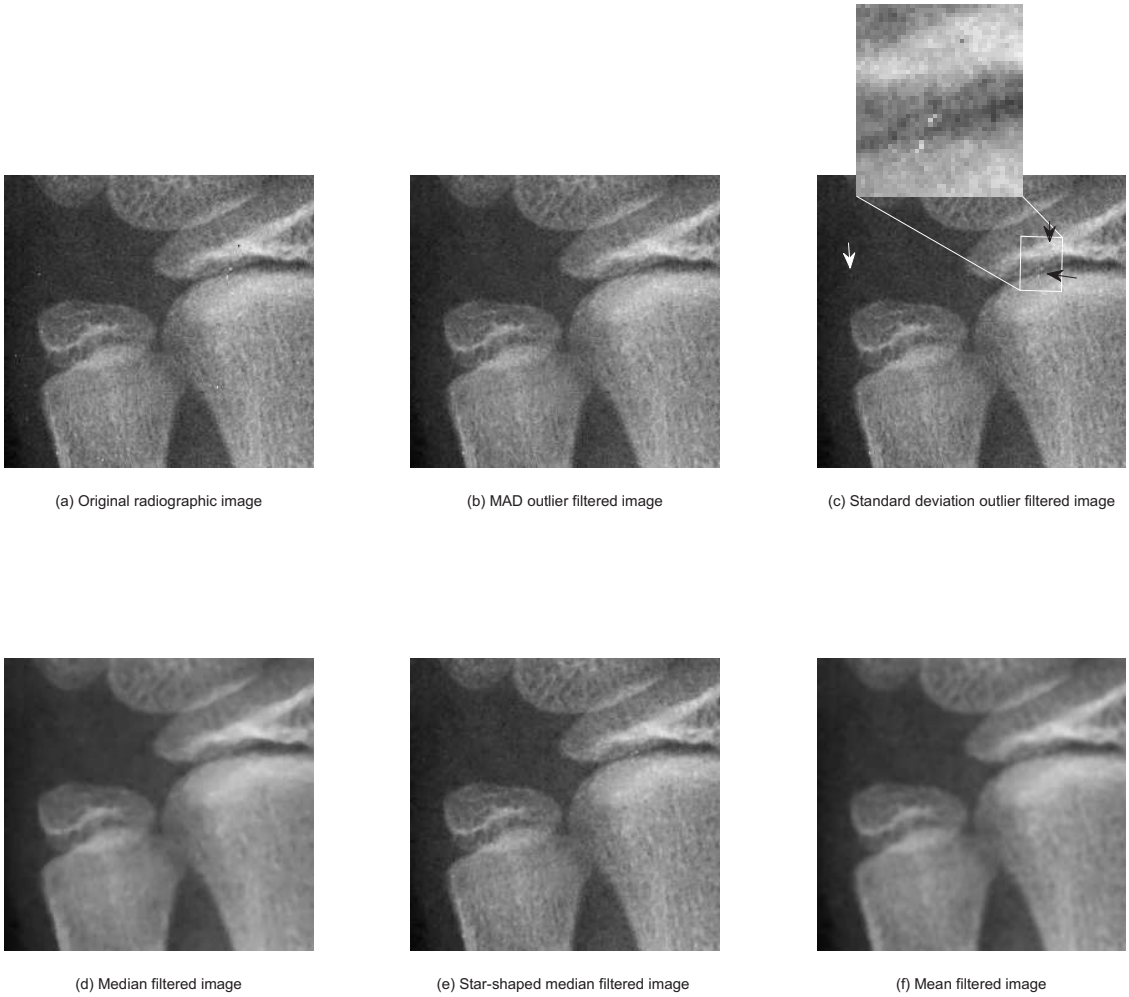


Figure 6.6 A comparison of the different outlier filters demonstrating the impact the filters have on the image noise and preservation of edges. There is both positive and negative impulse noise surrounding the radius and ulna bones in the original radiograph (a). The median, star-shaped median, and mean filter significantly alter the image noise (d)-(f), but the mean filter causes the most blurring because this filter is essentially a low-pass filter. The MAD-based outlier filter, (b), removes the majority of the impulse noise, with the exception of some residual impulse noise near the bone edges. The standard deviation based outlier filter has the same residual impulse noise as the MAD-based outlier filter in (b), but it has additional residual impulse noise (marked with arrows and illustrated in the inset). The impulse noise elimination capability of the star-shaped median filter seen in (e) is very similar to the MAD-based outlier filter, but there is qualitatively a significant change in the noise characteristics of the image when filtered with the star-shaped median filter. The MAD-based filter, star-shaped median filter, and standard deviation based filter all have difficulty removing impulse noise when it is present at sharp edges, as seen on the left-hand outer edge of the ulna bone (bone on the left-hand side of the image) and in the epiphyseal gap.

contrast, and structure between the reference and test images. The implementation in this section was based on the default parameters for the metric, as recommended in [Wang04]. To test the sensitivity of MSSIM to impulse noise, different frequencies of salt and pepper impulse noise were added to the image using the MATLAB® command `imnoise()`. Only the positive 'salt' noise was added to the clinical image. The sensitivity of the MSSIM index to impulse noise is illustrated in Figure 6.7. This demonstrates that the index has a good sensitivity for detecting impulse noise, and for the 0.1% impulse noise frequency used in this research the rate of change of the index is a maximum, suggesting a maximum sensitivity to impulse noise at low noise frequency.

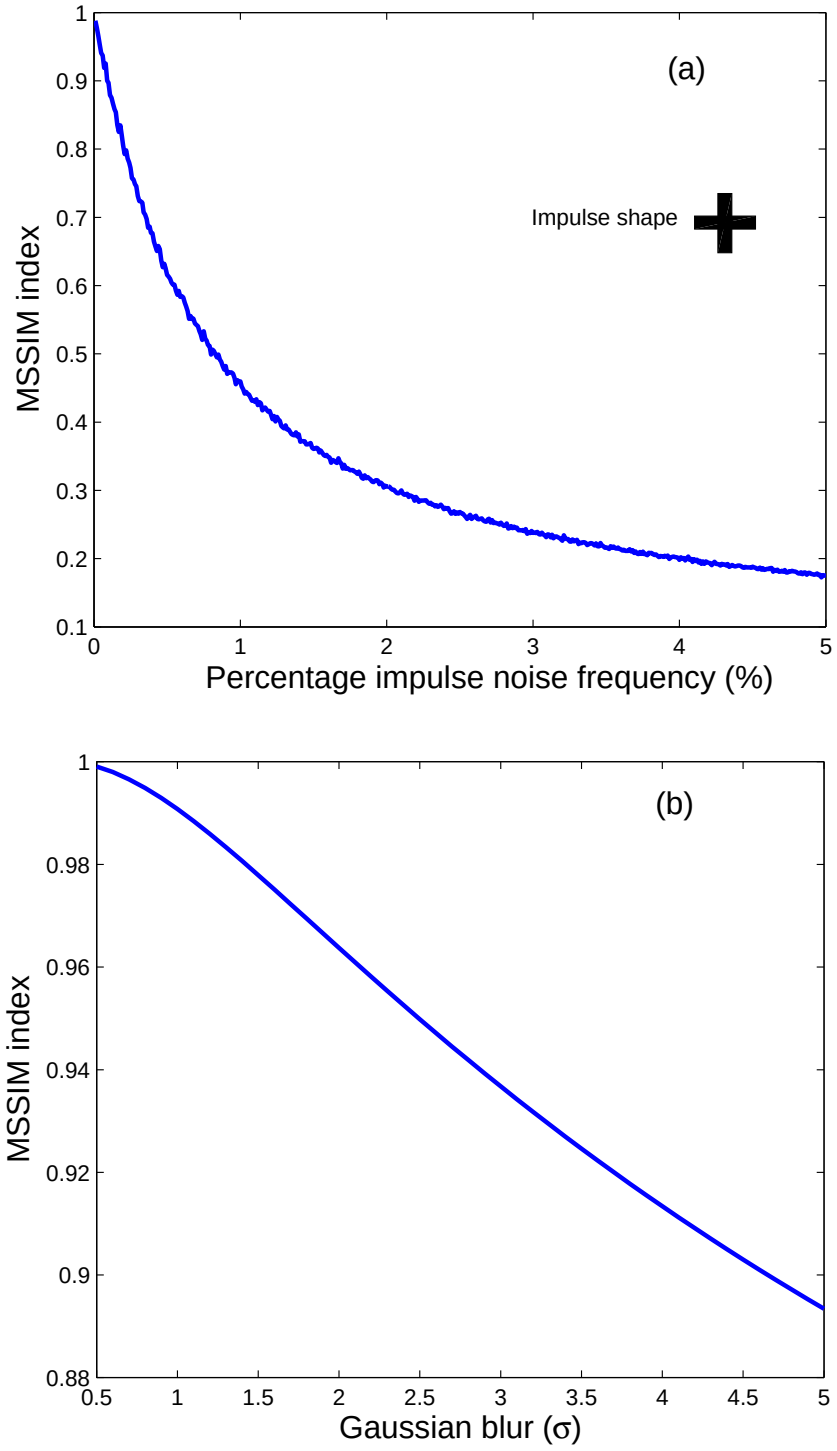


Figure 6.7 The mean Structure Similarity Index (MSSIM) for impulse noise and Gaussian blurring. (a) 3x3 '+' shaped impulse noise added to a clinical radiograph of the carpus. The MSSIM has good sensitivity to the impulse noise and has a high rate of change at 0.1% impulse noise, the frequency of noise used in these tests for comparison of the filters; (b) The MSSIM is less sensitive to blurring of the image using an isotropic Gaussian blur.

The response of the outlier filters to different types of impulse noise was measured by once again using an outlier pre-filtered clinical image to provide structure, but with minimal impulse noise to influence the results. Impulse noise with a frequency of 0.1% and a fixed amplitude of 3500 was added to the image (an intensity of 3500 being the maximum typical pixel intensity for bone). Four different types of impulses were added: single pixel, a 3x3 '+' shape, 3x3 block shape, and a 5x5 block shape. The results presented in Figure 6.8 illustrate that all three outlier filters performed well for the single pixel impulse noise. However, the star-shaped median filter and standard deviation filter performed poorly for the 3x3 block noise, as illustrated in Figure 6.9. The MSSIM results are presented in Table 6.1 for the image pre-filtered with MAD-based outlier detector. The results for pre-filtering using the standard deviation outlier filter have not been presented because they were essentially identical.

The MAD-based outlier filter results in Table 6.1 for the 5x5 pixel impulses (and 5x5 neighbourhood size) require some explanation because they reflect the robust nature of the median absolute deviation estimation. When the neighbourhood partially overlaps the impulse noise and contains less than 50% of the pixels of the impulse (50% being the breakdown value for the MAD estimator), then the MAD estimate will not be influenced by the impulse noise and the central pixel of the neighbourhood will be corrected if it is part of the impulse. As more overlap occurs, the MAD estimate breaks down and the pixel intensity is not corrected. This results in only partial correction of the impulse noise, as illustrated by an example in Figure 6.10. This partial correction of the impulse noise results in an increase of the MSSIM index, and in so doing, negates the value of the MSSIM index. Such large clusters of connected pixels were seldom observed in the Kodak CR 900 hand-wrist radiographs so larger neighbourhood sizes were not required.

Overall, the MAD-based outlier filter performs well for radiographs corrupted by impulse noise and this filter has been used in this research to filter the hand-wrist radiographs prior to further image processing.

(a) Original image with impulse noise removed

(b) Original image with added impulse noise

(c) Standard deviation−based outlier filtered image

(d) Star−shaped median filtered image

(e) MAD outlier filtered image

Figure 6.8 Comparison of the outlier filters with added single-pixel impulse noise of frequency 0.1%. The image in (a) had impulse noise removed then 0.1% single-pixel impulse noise added to produce the image in (b). All three outlier methods in (c)–(e) were effective in removing the single-pixel noise.

(a) Original image with impulse noise removed

(b) Original image with added impulse noise

(c) Standard deviation−based outlier filtered image

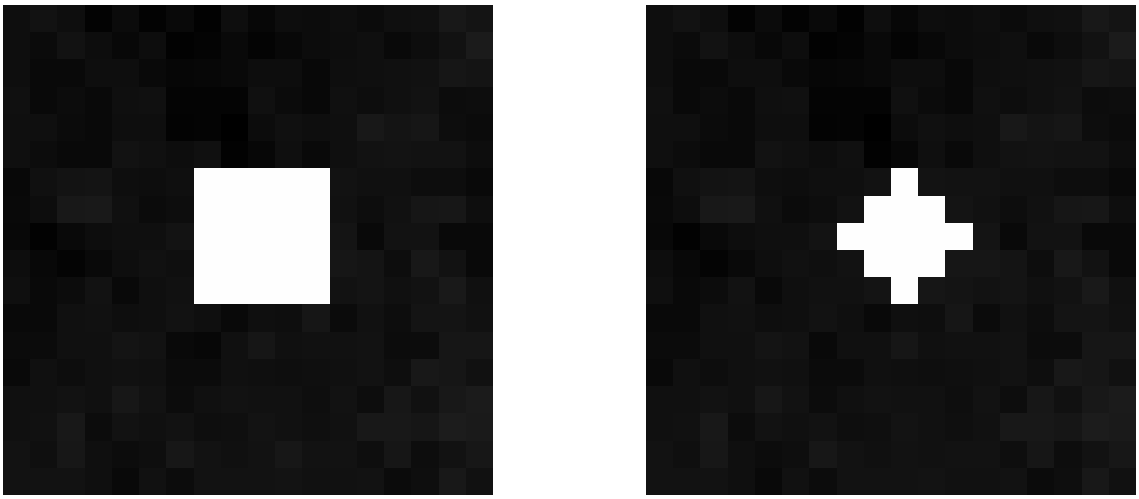
(d) Star−shaped median filtered image

(e) MAD outlier filtered image

Figure 6.9 Comparison of the outlier filters for added 3x3 block impulse noise of frequency 0.1%. Both the standard deviation and star-shaped filters (b) and (c) performed poorly for this type of noise. Although there are some obvious impulses remaining in the filtered radiograph, the results for the MAD-based outlier filter were much better than the results for the two other filters.

Table 6.1 Comparison of the outlier filters for different types of impulse added to a carpal image. A larger index value means the image is more like the uncorrupted, original image. The results for the impulse noise image are presented as a baseline for improvement. The MAD-based filter had the best performance for impulse noise elimination, within the uncertainties of the measurements.

Mean SSIM index	1 pulse	3x3 '+' noise	3x3 block noise	5x5 block noise
Impulse Noise image	0.986	0.891	0.829	0.709
Standard deviation filtered image	0.999	0.891	0.829	0.709
Star-shaped median filter	0.996	0.978	0.824	0.704
MAD-based outlier filter	0.999	0.999	0.994	0.769



(a) Original image with 5x5 pixel impulse noise

(b) MAD outlier filtered with 5x5 neighbourhood

Figure 6.10 MAD-based outlier filtering of 5x5 pixel impulse noise, (a), with a a 5x5 pixel neighbourhood. The robust nature of the MAD estimate means that pixels on the edge of the impulse are still calculated as outliers, but when more than 50% of the pixels in an analysis neighbourhood are impulse pixels, the standard deviation estimate using the MAD estimator breaks down and the impulse pixel intensities are not detected as being part of the impulse. This results in the star-shaped correction of (b).

6.3 Estimating signal-dependent radiographic noise

Image noise is often assumed to be an additive Gaussian white noise with zero mean and constant variance. Although this model of noise is useful when designing denoising algorithms, it is a poor approximation of radiographic noise because, as demonstrated in Section 4.3.1.2, the variance of the noise in radiographic images is signal-dependent, that is, non-stationary.

Although numerous methods have been developed for estimating the global stationary noise in single images (such as [Olse93]), such global statistics are useless for many real signals [Wyat07]. In recent years there has been growing interest in single-image methods of non-stationary noise estimation [Imme96] [Grav04] [Foi08]. These noise estimates have been used for both image denoising [Kuan85] and edge detection [Wyat07]. In this research the noise estimate is used as part of the parameter estimation for the exponential noise model of Eqn. (4.7). This signal-dependent noise model is then used in the anisotropic diffusion.

6.3.1 Noise estimation using the Kodak CR Exposure Index

One approach to noise estimation is to use the Kodak CR measurement called Exposure Index. The Exposure Index (EI) for hand-wrist radiographs is the average pixel intensity in the hand region of interest prior to applying the clinical processing[4]. It is calculated by the CR reader and stored in the DICOM file header. The average pixel intensity calculation means that the Exposure Index varies logarithmically with dose and, hence, can be used as a rough measure of radiographic exposure, as illustrated in Figure 6.11. If one avoids low dose radiographic exposures, then the EI value has a very consistent relationship with entrance dose to the CR plate for the hand-wrist phantom. The complication is that the EI values depend on a number of machine parameters and configuration settings.

The EI values for a CR plate depend on the energy composition of the x-ray beam and are therefore dependent on the kilovoltage setting for the x-ray generator and the equivalent filtration of the x-ray beam. This dependency may appear to diminish the usefulness of the EI as a reliable indicator of the dose to the CR plate, but what is required for the noise estimation model is not the dose to the CR plate, but the original pixel intensities before clinical processing. This is because these pixel intensities can be used to estimate the variance of the radiographic noise. A disadvantage with using the EI is that it is an average and only provides a single calibration value for the pixel intensities. By averaging the pixel intensities in the hand region of interest it is possible to find a scaling factor from this average

[4]The EI values are calculated using scaled outputs from the CR light collection system, so increasing dose produces increased EI values: for the raw image high EI values correspond to darker images, and lower pixel values in the final image.

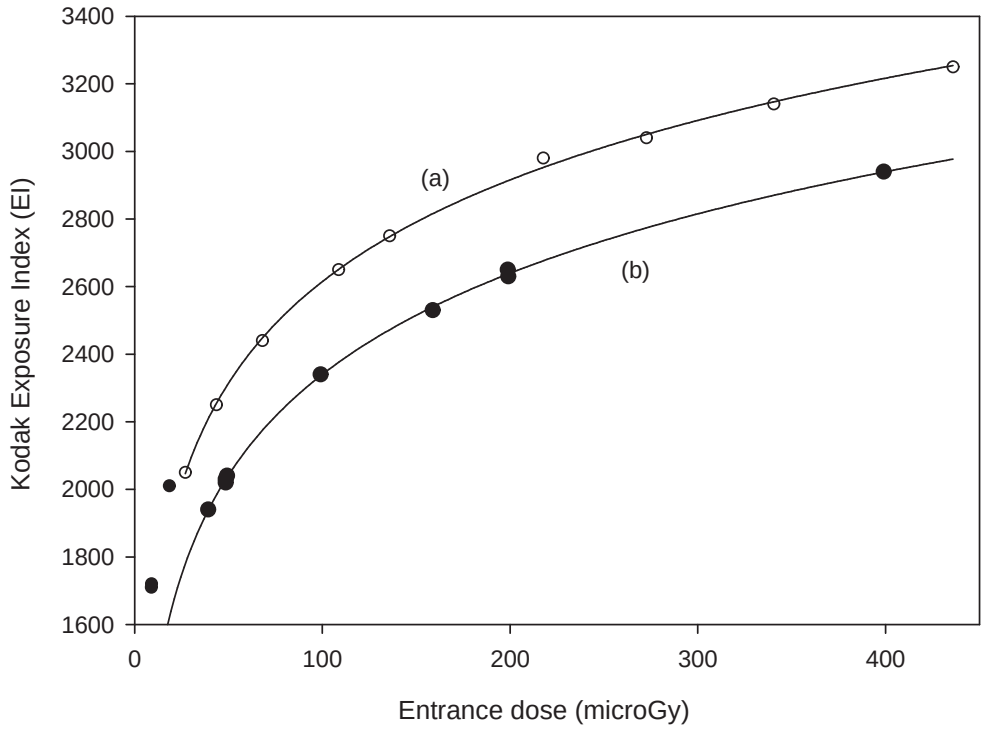


Figure 6.11 The variation in Kodak CR Exposure Index (EI) with entrance dose to the CR plate for a x-ray tube voltage of 52 kVp. (a) EI values for the stepwedge images in which there is no delineation of the anatomical region of interest, so the EI values are simply the average pixel intensity in the image. This explains the good logarithmic fit for the equation $EI = 1000\log(dose + a) + b$ with a correlation coefficient R^2=0.9993. (b) EI values for the hand phantom radiographs processed using the Kodak clinical protocol for hand radiographs. The outliers for doses below 20 μGy are most likely caused by the Kodak CR system calculating, in the presence of high noise levels, a hand region of interest that extends beyond the hand soft tissue outline and into the background pixel region. This would cause an increase in the measured EI value due to the high pixel intensities of the background. After removing doses below 20 μGy, the correlation coefficient for the logarithmic fit of $EI = 1000\log(dose + a) + b$, where a and b are constants, was R^2=0.9996.

to the EI. But the complication with use of the EI is that it is not possible to compensate for the contrast scaling of the image as part of the clinical processing, meaning that it would still be necessary to measure the noise content for the image. The EI can, however, be used for gross screening of the radiographs to reject the processing of the low dose images that do not fit the proposed noise model, as illustrated in Figure 6.11(b).

The EI values were collected for 97 hand-wrist radiographs that were used for bone age assessment. The histogram for the distribution of EI values is plotted in Figure 6.12. An EI increase of 300 corresponds to a doubling of the entrance dose to the CR plate. An EI value of 1900 has been chosen as the cut-off for the hand-wrist radiographs used in this study because of the affect the low dose optimisation has on image quality.

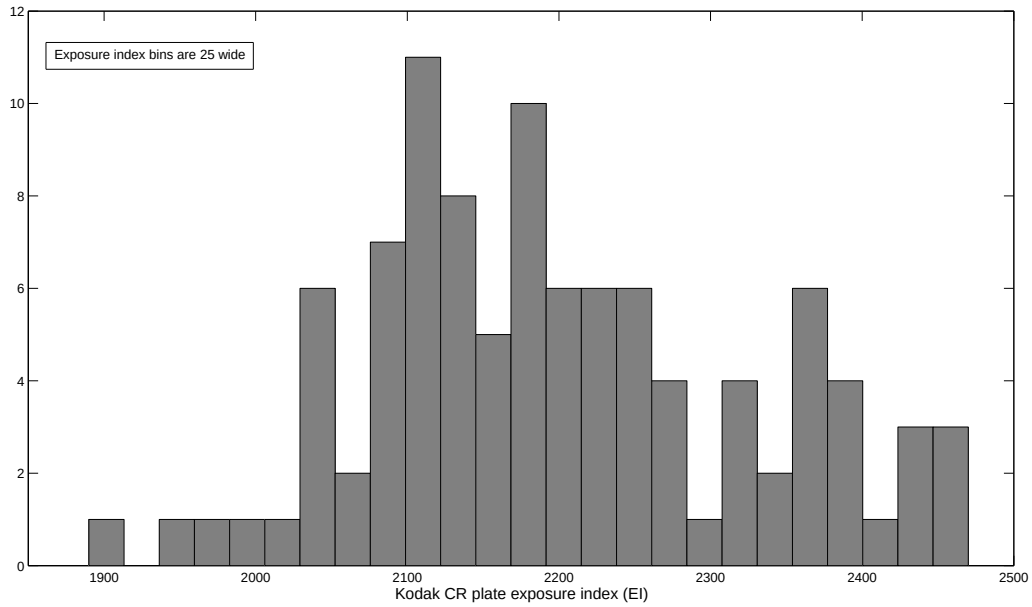


Figure 6.12 The range of Exposure Index (EI) values for 97 Kodak CR hand-wrist radiographs used for bone age assessment. The EI value of 1900 corresponds to a entrance dose of approximately 20 μGy on the hand phantom radiograph and is the lower limit of EI values used in this study. Below this EI value the low-dose optimisation of the CR system results in a noise model that is significantly different from Eqn. (4.7).

6.3.2 Noise estimation using local statistics

6.3.2.1 Background - noise estimation and local statistics

If one has access to multiple instances of the same image, it is possible to measure the signal-dependent noise on a pixel-by-pixel basis. In radiographic practice this is not possible. Apart from requiring multiple radiographic exposures, slight patient movement will cause significant changes in the calculated noise characteristics, especially near bone edges where the image gradients are high. Instead, it is necessary to estimate parameters for the noise model of Eqn. (4.7) using local image statistics from a single image.

When noise is estimated from multiple instances it is assumed that the only variation in pixel intensity is due to the random nature of the noise. In most situations the noise can be treated as an additive model of a signal plus a signal-dependent noise, where the signal-dependent noise is a distribution with zero mean value ($\mu = 0$) [Kuan85] [Foi08]. The variance of the pixel intensities for multiple instances of the same image allows one to estimate the noise variance per mean pixel intensity. With single images this variance calculation cannot be performed and it is necessary to measure the signal-dependent variance using local statistics.

If the image contained no structure, estimation of the noise would be a trivial exercise if

the pixel intensities are uncorrelated. To estimate the noise it is necessary remove the influence of the image structure and a number of methods have been used for this. The first and most common method is to apply spatial smoothing of the image to remove as much noise as possible, then subtract the smoothed image from the original image to give a noise image [Grav04]. This makes the noise estimation very dependent upon the image smoothing and if the smoothing of edges occurs, then the resulting noise image will still contain components of the structure. To a certain extent the influence of the edge structure in the noise image can be reduced by masking-out edge components [Grav04] [Olse93] [Tai08] [Corn03]. Alternatively, the noise component of the image can be estimated by also treating the noise as having high spatial frequency components and isolating these components from the rest of the image using high-pass filters. Another way of thinking of this is to treat the structure as persisting across multiple scales and to employ wavelet methods to find the fine detail components of the image, this being the noise [Foi08] [Hein06]. Once again, the influence of edges needs to be removed from the detail (noise) image and Foi *et al.* use an interesting technique of thresholding the detail image using the sum of the moduli of gradient and Laplacian operations. They describe this as a 'heuristic way to obtain "thickened" edges' that are excluded from the noise analysis [Foi08]. An argument against using wavelet methods is that they can cause erosion of highly curved parts of the image because the wavelet basis functions are not steered relative to structure components of the image [Wyat07]. A proposed alternative is to steer filters in directions orthogonal to image structures where the variance will be dominated by noise rather than image structures [Wyat07]. Either way, once suitable noise pixels have been found it is possible to calculate mean-standard deviation pairs to give the relationship between the mean pixel intensity and the noise associated with the pixel intensity.

Asymmetry of the pixel intensity distribution

A complication with noise estimation based on an additive model is that although this applies to the Poisson noise arising from exposure of the CR plate, the nonlinear nature of the logarithmic conversion means that the additive model becomes difficult to work with. This can be studied by modelling the Poisson noise using a Gaussian distribution and performing a transformation of random variable calculation from a Gaussian probability density function using the transformation $I_{mean} + I_{noise} \rightarrow log(I_{mean} + I_{noise})$ (see, for example, [Jahn02, p81] for details of random variable transformation). Although this transformed probability density function can be derived analytically, it can also be demonstrated using a simulation of the Poisson noise and performing the logarithmic conversion. The MATLAB® random number generator `randn()` was used to generate a sample of 10^6 values. The Poisson noise was modelled using a continuous Gaussian distribution approximation. This was necessary so that large values of λ could be used in the simulation. Large values of λ could not be used with the Poisson random number generator utilised in Chap-

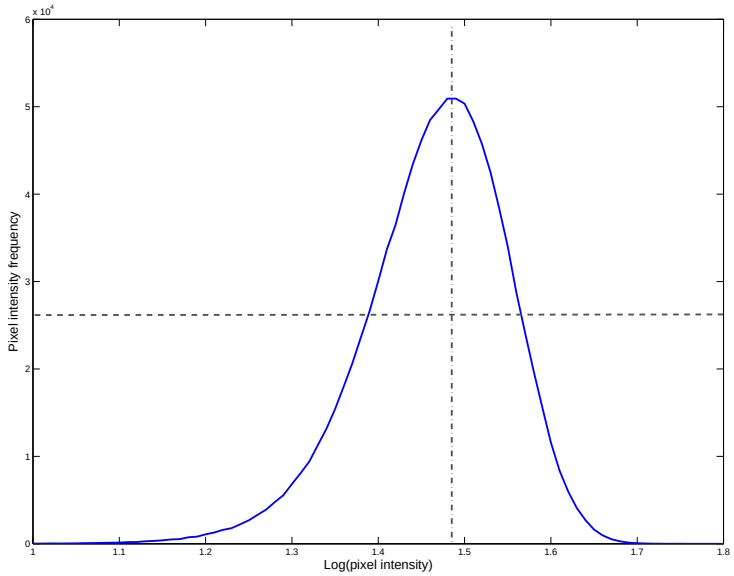


Figure 6.13 Simulation of the logarithm of a Poisson distribution using $\lambda = 30$ and 1×10^6 simulated samples. The asymmetry in the distribution was more pronounced for low values of λ.

ter 4. The results for the simulation are shown in Figure 6.13 for $\lambda = 30$. The distribution of pixel intensities is clearly asymmetric. This asymmetry decreased as λ was increased. The use of the Gaussian approximation meant that the non-transformed random variable was symmetric for small values of λ. Therefore, the asymmetry present in the curve of Figure 6.13 is due to the logarithmic transformation not the underlying approximation of the Poisson distribution.

The impact of this asymmetry is demonstrated nicely by the Anderson-Darling test for the normality of a sample [Ande54] [DAgo86]. This test is one of nearly 40 different normality tests that have been developed (as cited by [Yazi07]). The Anderson-Darling test was chosen instead of the popular Kolmogorov-Smirnov (K-S) test because of the good performance of this statistic, a performance that is relatively independent of sample size [Yazi07], especially once a small sample correction factor has been applied [DAgo86]. The Anderson-Darling test should normally only be applied to symmetric distributions because it is sensitive to asymmetry. This sensitivity has been exploited to test for asymmetry. The critical value for the test statistic was calculated for a $p = 0.05$ significance level and for the case of an unknown mean and variance of the sample [DAgo86] (the critical value was 0.752). The test for normality was applied to simulated 100x100 pixel images that had signal-dependent Gaussian distributed noise. The test was applied to both the raw simulated image and a logarithmic transformation of the image data – this simulated the CR data conversion. The Anderson-Darling statistic was calculated for 7x7 neighbourhoods of pixels. If the statistic for the neighbourhood exceeded the critical value, the pixel at the centre of the neighbourhood was marked as a one; otherwise, it was marked as a zero.

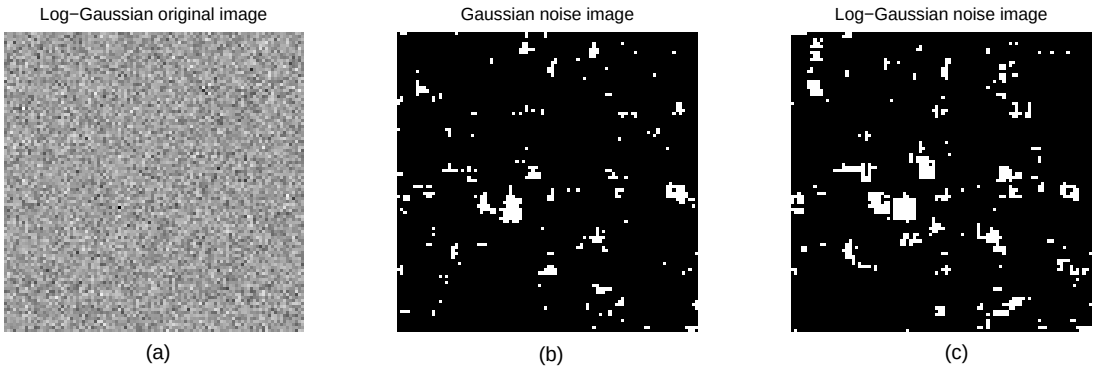


Figure 6.14 The Anderson-Darling test for normality applied to simulated images with signal-dependent Gaussian distributed noise. This series of images was for a value of $\lambda = 200$ for the Poisson distribution that the simulation was approximating. (a) Logarithmic transformation of the simulated image using $y = 1000\log(x)+10$; (b)-(c) binary images corresponding to neighbourhoods in which the distribution of pixel intensities were not Gaussian distributed (at p=0.05).

This created a binary image in which values of one corresponded to the neighbourhoods that exceed the critical value. The result for one simulated image is shown in Figure 6.14. The distribution of pixel intensities was tested for a range of mean pixel intensities, corresponding to different λ for the Poisson distribution. As the value of λ decreased more pixel neighbourhoods failed the normality test for the logarithmic transformed image data. The Gaussian distributed image data varied slightly around 0.05 (corresponding to p=0.05 for the critical value), as illustrated in Figure 6.15. The Anderson-Darling test applied to a clinical hand-wrist radiograph, Figure 6.16, demonstrates the interesting result that apart from the bones edges, the pixel intensity distribution for the bones is more Gaussian-like than the soft tissue and background. This is probably due to the clinical processing protocol influencing the non-bone noise. This could explain the findings of Section 4.3.4 for noise in the hand phantom radiographs, especially why the noise levels in the background region could not be modelled using Eqn. (4.7). Furthermore, with a large number of pixel neighbourhoods in the background region failing the Anderson-Darling test for normality, this explains why the MAD-based outlier filter for removing impulse noise had a larger number of false detections than expected for a Gaussian distribution of pixel intensities, as discussed in Section 6.2.2.1.

Although there is obvious asymmetry in the distribution of pixel intensities for low mean values (small λ), the question that remains is whether such values of intensities reflect the situation for clinical hand-wrist radiographs. This can be checked by considering the detection of photons. Ignoring the influence of the polychromatic x-ray beam upon the detection of photons, the likely number of photons that could be detected per pixel was calculated using the x-ray spectrum simulation program Spectra (National Radiation Laboratory, New Zealand). For 52 kVp and 1 mAs x-ray exposure, the radiation dose transmitted through

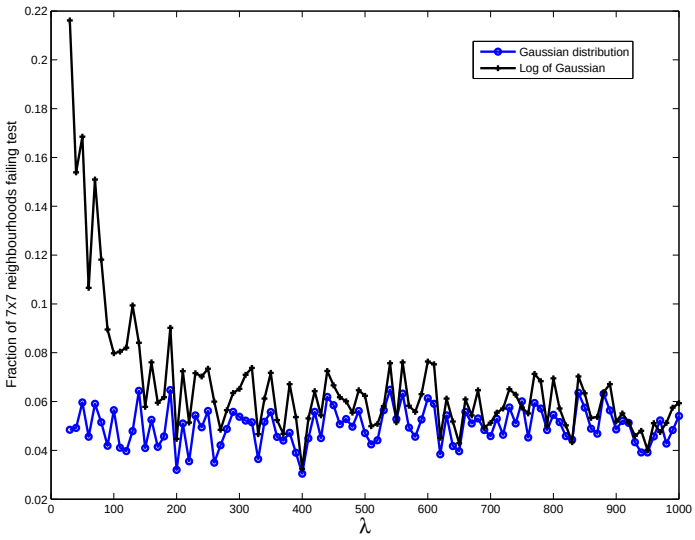


Figure 6.15 Differences in failure rates for the Anderson-Darling normality test when applied to images with Gaussian distributed noise and log-Gaussian distributed noise. The fraction of neighbourhoods failing the test is the fraction of 7x7 pixel neighbourhood statistics exceeding the test critical value, compared with the number of pixels in the image. The average value of 0.05 for the Gaussian distributed noise indicates that the Anderson-Darling statistic is performing correctly because one would expect a fraction of 0.05 to fail for a critical value based on p=0.05. The increased failure rate of the log-Gaussian distribution for low values of λ can be attributed to the asymmetry in the pixel intensity distribution.

20 mm of soft tissue and 20 mm of cortical bone would be approximately 0.56 μGy. This is a very low entrance dose to the CR plate. For this dose the flux of photons incident upon the detector is approximately 17000 photons/mm^2. The Kodak CR plates have a pixel size of 0.097x0.097 mm^2, so the average number of photons incident upon the plate per pixel is at least 160. This does not allow for the detection efficiency of the detector, but even for only 50% detection efficiency, only the highly attenuated areas of the radiographs would have an associated asymmetric pixel intensity distribution. This is partly supported by the results presented in Figure 6.16 which show pixel neighbourhoods that fail the Anderson-Darling test along some parts of the highly attenuating edges of the phalanges where there is compact bone. However, negating this evidence is that the soft tissue-to-air boundary is also clearly visible, suggesting that the pixel neighbourhoods fail at the edges because of the pixel intensity contrast in the neighbourhoods.

The bone and soft tissue edges that are visible in Figure 6.16 suggest that local statistics could be used for segmentation and the finding of contours of the bones, and connected component analysis may help find the soft tissue contour. However, the contours for the bones are much more broken than those for the soft tissue, and in particular the carpal bones are barely visible in this image that is typical for clinical hand-wrist radiographs. It

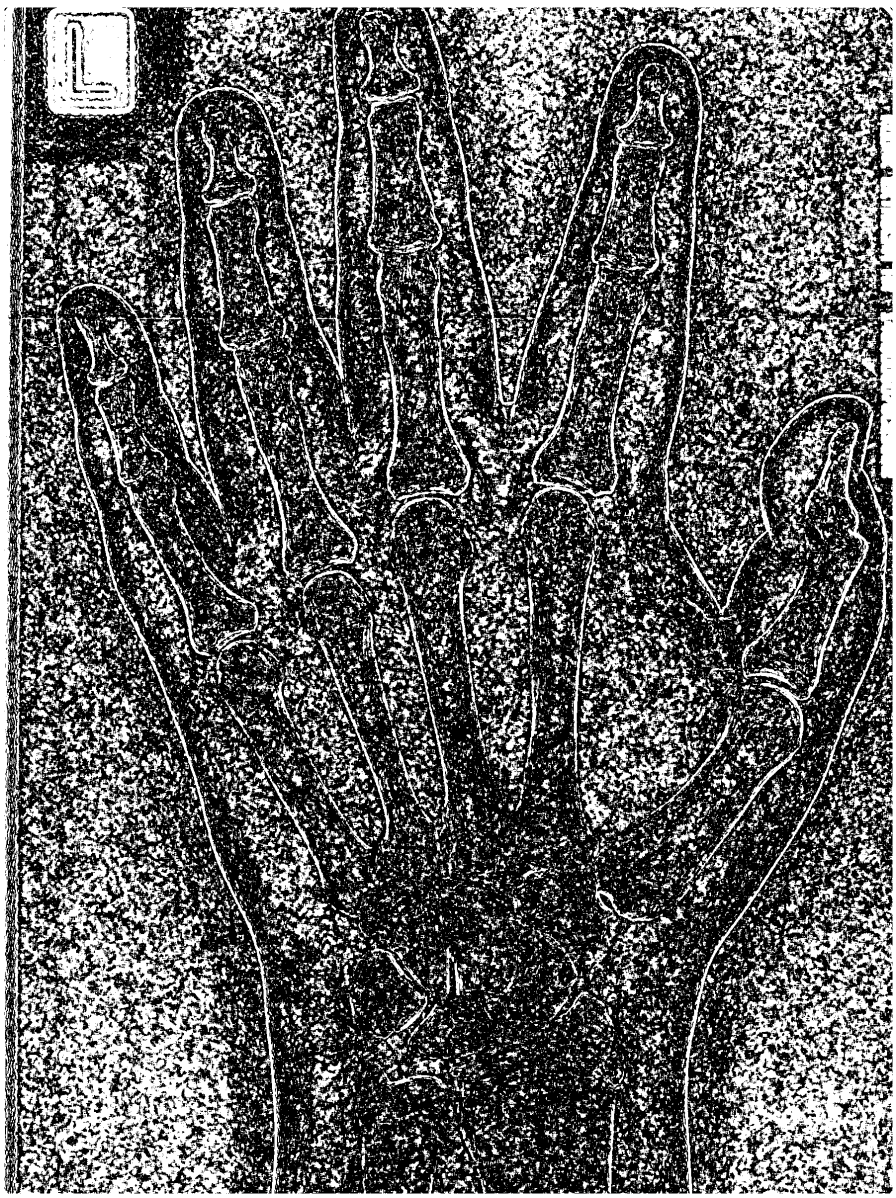


Figure 6.16 The Anderson-Darling normality test applied to the clinical hand-wrist radiograph of Figure 6.2(a) using 7x7 neighbourhoods and a critical value based on p=0.05. The white pixels correspond to failure of the normality test for the neighbourhood centred at the pixel. Approximately 34% of the pixel intensities are labelled as not being from a Gaussian distribution, with a 0.05 probability of this labelling being incorrect.

does demonstrate the idea of edges being outliers of a statistical distribution. This explains why in Section 6.2 the MAD-based outlier filter was less successful at removing impulses on the edges of bones because the dispersion of pixel intensities was so large that it was difficult to find statistically significant impulses above the pixel intensity variations in the analysis neighbourhood.

6.3.2.2 Noise estimation using the MAD robust statistic

The measurement of noise in the hand phantom radiographs of Section 4.3.4 was made possible by designing the phantom to have homogeneous areas and by only measuring the noise in these areas. Such large areas of homogeneity are not present in the clinical radiographs, except in the background region and analysis of the background is excluded from the noise estimation because it is not of interest for denoising (it is outside of the hand contour). The conundrum is that although homogeneous regions can be used for noise estimation under the assumption that they are not affected by image structure, truly homogeneous regions will contain no noise, invalidating the noise estimation. The noise methods discussed earlier approach this problem by estimating the additive component of the noise in regions that are away from image structures so that the image structure does not influence the variance measurement. However, the purpose of developing and measuring the signal-dependent noise model is to use the model for image denoising. The objective is to find the parameters of the noise model then apply this formulation on a pixel-by-pixel or region basis in the denoising algorithm. This requires the model to work for every pixel within the hand soft tissue outline, independent of structures such as bone. In this situation it is not a sensible approach to only find candidate regions and model the noise using these because the noise estimation is likely to breakdown for the areas that have not been included. In particular, it does not make sense to estimate the local variance of a pixel neighbourhood because this will be significantly influenced by local structure. The method adopted was to estimate the mean intensity for the pixels, bin the pixels according to this intensity, and use the MAD robust estimator to find the standard deviation for the resulting sample of pixel intensities. This produces a collection of mean–standard deviation pairs that can be used to estimate the parameters of the noise model. This approach is similar to that by Gravel *et al.* [Grav04] and Foi *et al.* [Foi08], including the use of the MAD robust estimator. However, for the application at hand both Gravel *et al.* and Foi *et al.* used methods to avoid the influence of image structures. Additionally, Gravel *et al.* estimated the noise component by subtracting a smoother version of the image from the original, whereas Foi *et al.* used wavelet details to find the noise component.

Whereas in the impulse outlier detector the MAD estimator is applied to local neighbourhoods, in this application the estimator is applied to a collection of mean–standard deviation pairs from across the image. The success of this approach is dependent upon the

correct estimate of the mean pixel intensity, and measurement of the standard deviation for the collection of pixels associated with the mean pixel intensity. Both of these are now investigated.

Evaluation of the MAD robust estimator for measuring noise statistics
Although the MAD is a robust estimator, it suffers from three major drawbacks: sensitivity to asymmetric distributions [Rous93], increased variance in estimates compared with the standard deviation [Foi08], and it is asymptotically biased to overestimate the variance of a sample [Foi08]. The main concern was with the effect of the asymmetry because as presented earlier, the simulations indicated that the logarithmic conversion of the CR data results in an asymmetric distribution for low pixel intensities. This effect was investigated by simulating 10,000 pixel intensity samples from a Poisson distribution where the expected pixel intensity is the value of λ for the Poisson distribution. Because the investigation focussed on small values of λ for the Poisson distribution, a Poisson random number generator was used [Mart02, p102]. The simulation method was first checked before applying it to the logarithmic transformed data. The results presented in Figure 6.17 were as predicted for the sample mean and variance, but the MAD estimator had unexpected discrete steps in the estimates of sample standard deviation. The sample standard deviations were as expected so the MAD estimator was thought to be the cause, not the random number generator. This was confirmed by the results presented in Figure 6.18 where the continuous Gaussian distribution was used to simulate the Poisson distribution, and simulations were performed with and without rounding the simulation values to integers – with the rounding used to simulate the detection of photons. The rounding of the values to integers has a large impact on the MAD-based standard deviation estimates.

The number of x-ray photons detected per pixel is an integer number and to a reasonable approximation the stimulated light output of the CR plate for any individual pixel will be an integer N times a conversion factor s. When this detected signal undergoes a logarithmic conversion, the result is $\log(sN) = \log(s) + \log(N)$. This means that it does not matter at which stage prior to performing the logarithmic conversion the rounding of integers is applied because one can always find a new s to scale the conversion. However, it is important that the final pixel intensities calculated are rounded to integers because this is the case with the output from the CR reader. Using this approach, the MAD estimator was tested for wide range of simulated CR pixel intensities. The results presented in Figure 6.19 for the MAD estimator were different than those presented in Figure 6.18 in that the significance of the steps in the MAD estimates was reduced, but not eliminated, by the logarithmic operation. Furthermore, for the logarithmic converted data, the MAD estimate of standard deviation was in good agreement with the sample standard deviation up until a variance of approximately 3000, corresponding to a pixel intensity of 2300 and $\lambda = 63$. These pixel intensities correspond to low x-ray exposures or noisy regions of radiographic

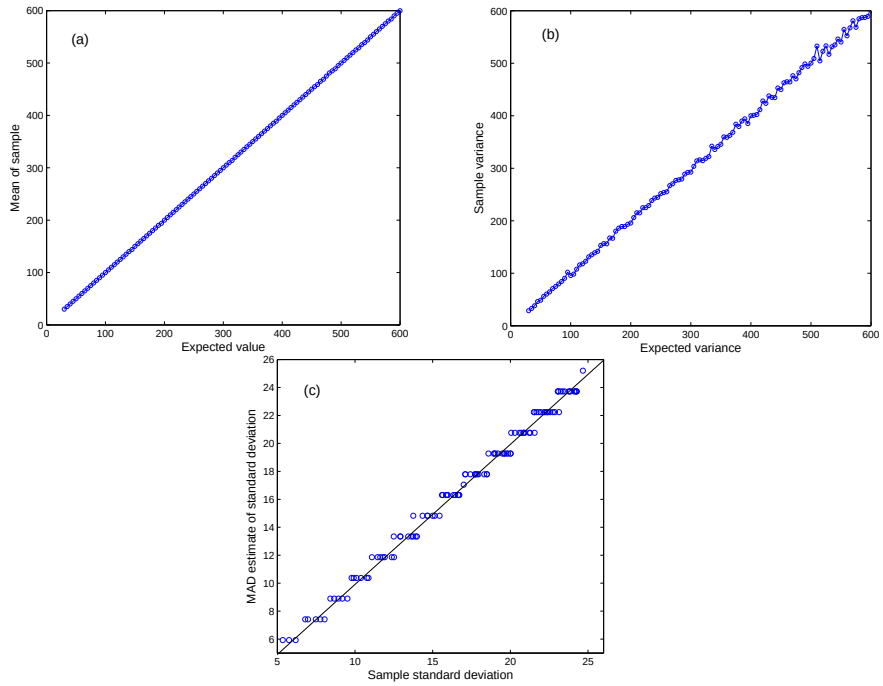


Figure 6.17 Evaluation of the MAD robust estimator for Poisson distributed noise. The simulation produced 10,000 samples for values of $\lambda \in [100, 600]$. (a) Mean of the sample; (b) Sample variance versus the expected same variance, in this case the value of λ; (c) MAD-based estimation of sample variance. The discrete steps were caused by the discrete integer values of the Poisson distribution.

images. What is especially important to note from these results is the good agreement between the MAD estimates of sample standard deviation and the standard deviation predicted from the noise model of Eqn. (4.6). There was less agreement between the sample variance and the expected variance for high variance values (variance values higher than those observed for the hand phantom results of Figure 4.11). This most likely reflects the breakdown assumptions of the noise model for low pixel intensities (asymmetry for small λ), in which case the very good agreement between the MAD estimate and the expected standard deviation is simply coincidental. These simulations used 10,000 sample pixel neighbourhood to evaluate the MAD estimator. The size of this sample was decreased to 25 pixels and the simulations repeated. Such a small sample is very challenging for many estimators, yet from Figure 6.20 it can been seen that the MAD estimator produced reasonable estimates of the sample standard deviation. This result supports the use of 5x5 pixel neighbourhoods employed in the outlier detector discussed in Section 6.2.2.1. The spread of the estimates reflects the known drawback of the increased variance for the MAD-based estimates compared with the standard deviation [Foi08].

Finding the mean pixel intensity map and estimating noise

A mean pixel intensity map was used to bin the pixel intensities. This is an image of the

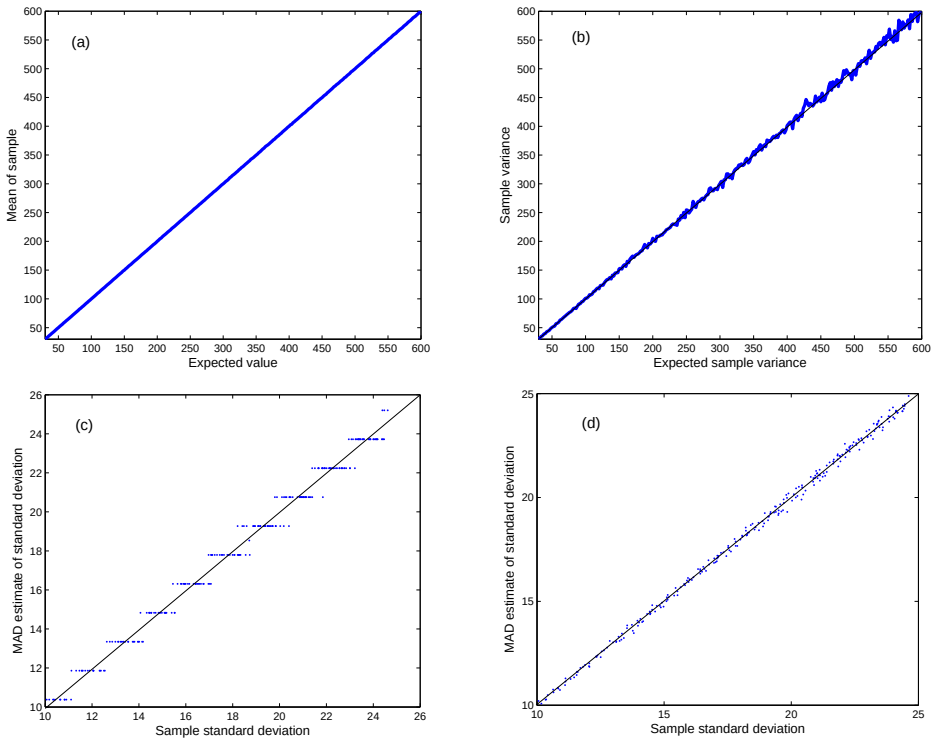


Figure 6.18 Evaluation of the MAD robust estimator using a continuous approximation of the Poisson distribution using 10,000 samples and $\lambda \in [100, 600]$. (a) Mean of the samples; (b) Sample standard deviation – as expected because for a Poisson distribution $\sigma^2 = \mu$; (c) MAD estimate of standard deviation for rounded simulation values. (d) MAD estimates for no rounding of the simulation values. It is clear that rounding to integer values has a significant effect on the MAD estimation.

mean pixel intensities of a neighbourhood. Unlike the methods by Gravel *et al.* [Grav04] and Foi *et al.* [Foi08], the intention was not to avoid image structures, but rather to make a robust estimate of the mean pixel intensity within the structure, and to include these pixel intensities in the standard deviation calculation. This approach was taken because to use the noise model in the denoising routine it was important to calculate it on a pixel-by-pixel basis; hence, it did not make sense to exclude edges. Finding the image of mean pixel intensities is, therefore, an interesting challenge because what is required is effectively a denoised image, but with minimal blurring of the edges because this would affect the edge pixel statistics. This is a paradoxical situation where to estimate the noise directly from collections of pixels, one requires no noise in the pixels to estimate the mean. This lends itself to an iterative process, but iteration was not adopted in this research because the advantage of a more accurate noise estimate was not considered necessary for the denoising application. (This may be worth considering for future research.)

There was no consistency in the estimate of the noise if the image was not filtered prior

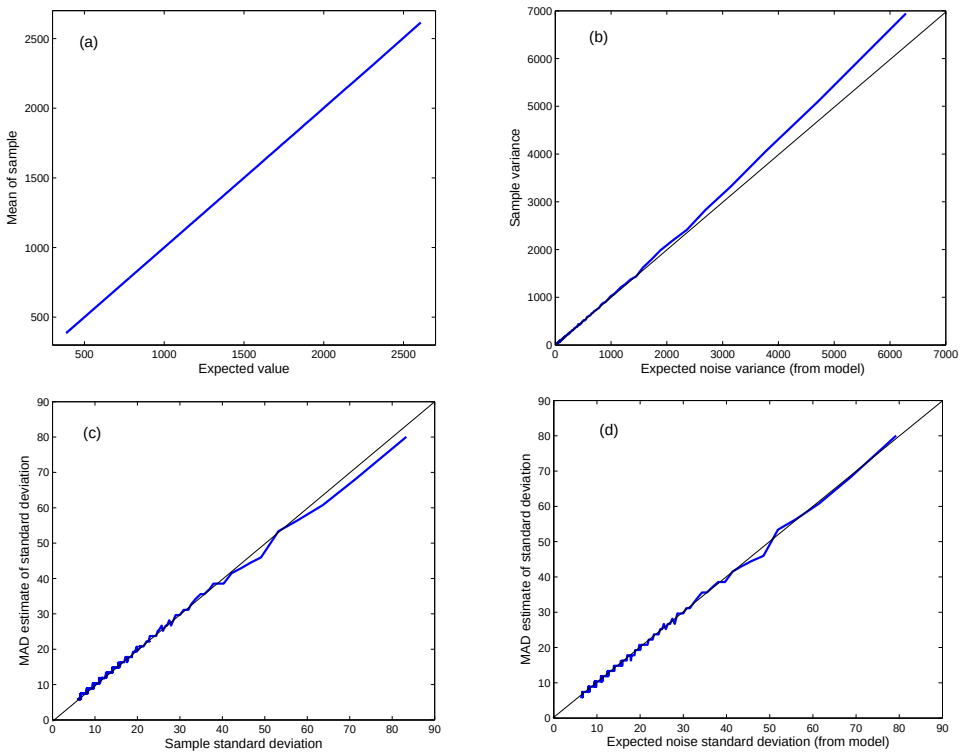


Figure 6.19 Simulation of Poisson noise to test sensitivity of the MAD estimator to logarithmic converted CR data with an asymmetric distribution for small values of $\lambda \in [30, 5000]$. The logarithmic conversion was of the form $y = 4095 - 1000 \cdot \log(x) + 10$ to simulate CR data. Although there are still discrete steps in the MAD estimates, there remains good agreement between the MAD estimate and the expected standard deviation (d). There appears to be a breakdown of the noise model for variances above approximately 3000, as apparent from the graph of (b) and (c). This corresponds to pixel intensities of more than 2300 or $\lambda \lesssim 63$. For this small value of λ the pixel intensity distribution becomes asymmetric and this could be another possible explanation for the apparent breakdown of the model.

to performing the noise estimation. This filtering reduced the influence of the noise on the mean pixel intensity estimates. This clearly shows the importance of the filter. Two filters were investigated: the median filter, and the isotropic Gaussian blur. The image was firstly filtered to produce an estimate of the mean pixel intensities, then masked using a soft tissue outline of the hand that was found using the method described in Section 5.2.4. The masking was for consistency in the noise estimation because the Kodak CR system adjusts the image characteristics based on the pixel intensity contents of the hand soft tissue outline. The resulting mean pixel intensity image was binned into 350 pixel intensity bins each 10 wide, starting at a pixel intensity of 200 and finishing at 3700. The pixel intensities were not of interest above and below these limits, as discussed in Section 5.3. The bin size was found empirically based on a trade-off between being too wide and changing the standard deviation estimates, and being too narrow and causing excessive variance in the estimate.

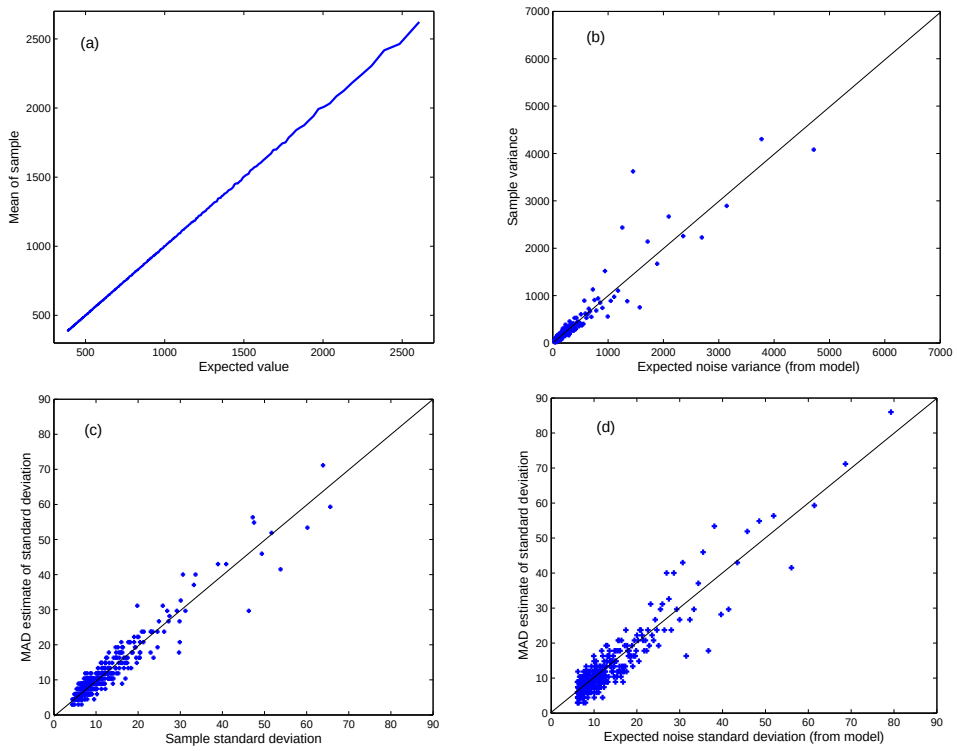


Figure 6.20 Evaluation of the MAD standard deviation estimates for CR data and for a small 25 pixel neighbourhood. The MAD estimate (c) has larger spread of data points than the standard deviation (b), but the MAD operation still produces a reasonable estimate of the standard deviation for such a small sample size.

The MAD robust estimator was used to estimate the standard deviation σ_k of the pixel intensities in each bin k. For each bin k the pixel intensities should be approximately Gaussian distributed, so the mean pixel intensity $\bar{I}_k$ was estimated using the median as a robust estimator of central tendency. The variance of the standard deviation estimate was also calculated for each bin k using [Foi08]

$$\mathrm{Var}\{\sigma_k\} = {\sigma_k}^2 \cdot d_k^{mad}, \quad d_k^{mad} \simeq \frac{1.35}{n_k + 1.5} \tag{6.3}$$

where n_k is the number of pixels in bin k (Gravel *et al.* [Grav04] present a different formula for the MAD variance calculation, but the two agree within 4%).

The effect of the filter on the noise estimation is illustrated in Figure 6.21 for a typical hand-wrist radiograph. All filters resulted in an approximately linear relationship between $log(\sigma_k)$ and the mean pixel intensity $\bar{I}_k$. However, as the amount of spatial spread increased for the Gaussian blur filter, the noise estimation dropped off for high pixel intensities and increased for low pixel intensities. When the amount of filtering was very low, as with the isotropic Gaussian blur filter with $\sigma = 0.5$, the noise is underestimated. As the blurring

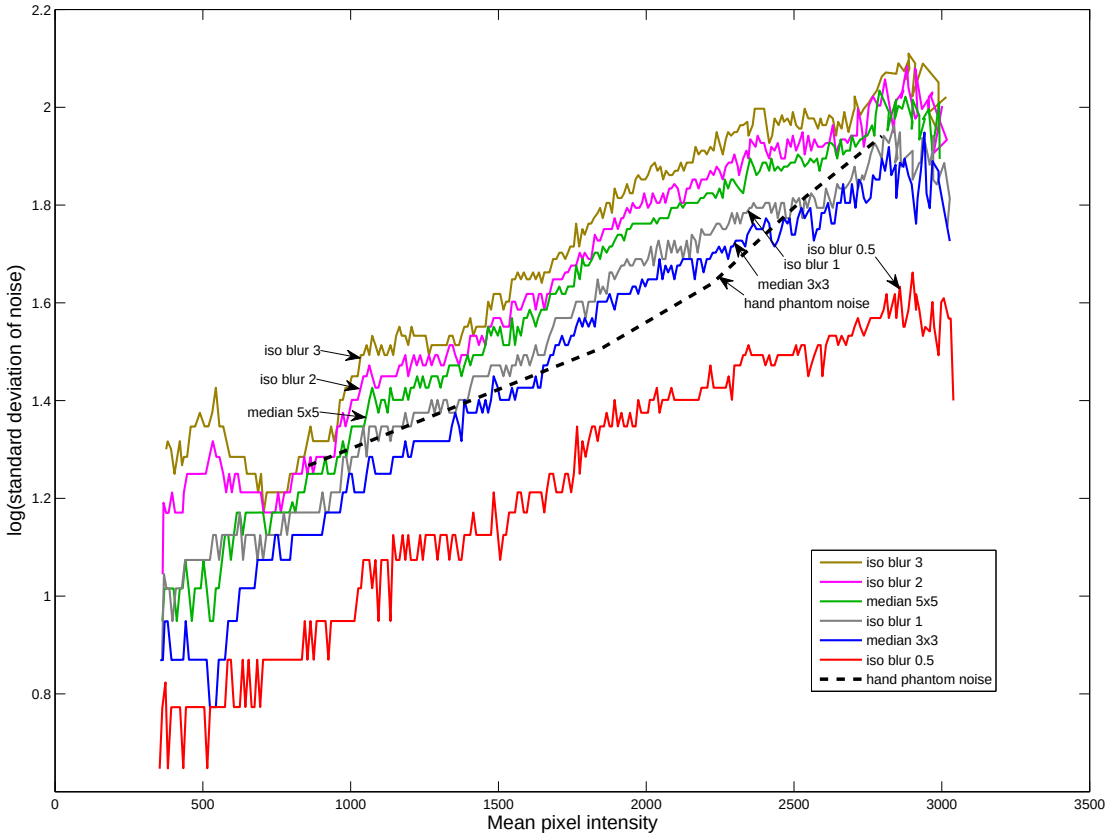


Figure 6.21 The effect of different mean pixel intensity estimators on noise estimation for a typical radiograph, in this case that of a of a young adult 16 years of age. Six filter configurations were tested: 3x3 and 5x5 median filter; isotropic Gaussian blur filter (iso blur) with spatial spread $\sigma = \{0.5, 1, 2, 3\}$. The slopes of the curves are very close to those for the hand phantom in Figure 4.11 The dashed plot is the hand phantom noise derived from homogeneous regions of the hand phantom radiograph for an entrance dose of 99 μGy, slightly higher than the predicted entrance dose of 88 μGy for the clinical radiograph. The 3x3 median filter was selected for noise estimation. (Please note that the original figure has coloured lines)

is decreased further towards no filtering at all, the noise estimation becomes very erratic. This behaviour can be explained by the mean pixel intensity being incorrect, so the pixel intensities are spread across a number of bins not related to the mean of the distribution of pixel values. In the limiting case of no filtering, the pixel intensities are assigned to bins relating directly to the value of the pixel, so all pixels in the bin will have very similar intensities and the variance will decrease to zero. The hand phantom noise values for a 99 μGy entrance dose were added to Figure 6.21 for comparison. These hand phantom results were chosen by converting the EI value for the clinical radiograph to an equivalent entrance dose of 88 μGy using the calibration of EI to dose presented in Figure 6.11. The 3x3 median filter was chosen for this research because of the robustness of the median calculation, and because it had similar performance to the Gaussian blur with $\sigma = 1$.

In accordance with the noise model Eqn. (4.7), linear regression with equal weights was performed on the mean-standard deviation pairs after taking the logarithm of the standard deviation for each bin. Although more advanced curve fitting can be employed [Foi08] [Grav04], the elementary curve fitting by linear regression gave surprisingly good results. This is probably because of the influence of the logarithmic function on outliers. To improve the robustness of the noise estimate, only bins with more than 100 pixel samples were included in the linear regression analysis. Figure 6.22 shows, on a linear scale, the results of the linear regression for a typical radiograph. It is evident that there is reasonable agreement with the exponential noise model, but there is poor agreement for high intensity pixels of 2500 or more. For the higher intensity pixels there are fewer samples in the bins and the errors in the estimates increase, as illustrated in Figure 6.23. The noise underestimation for high intensity pixels also can be explained by the high intensity edges of bones being incorrectly binned because the median filter could not denoise the bone edges correctly.

The results are different for the images of younger children. The wrist, for example, contains fewer bones, so it can be difficult to collect statistics about the bones. An example is shown in Figure 6.24 and demonstrates that although there is a significant amount of variability in the fitting, overall there is a trend for the noise to increase exponentially as the pixel intensity increases.

As a final check of the noise estimation method, the pixel values of the TG18-BR test pattern [Same05, p45] were used as mean pixel intensities for the simulation of an x-ray of the pattern. The TG18-BR pattern contains varying size and contrast checkerboard objects (see the inset of Figure 6.25). The pixel values of the 1024x1024 pixel test pattern were used as λ values for simulation of Poisson distributed photon counts. The photon counts per pixel C_i were converted to CR pixel intensities P_i using the logarithm conversion $P_i = 1000 \log(C_i) + 10$. This gave pixel intensities comparable to those for the CR images. The

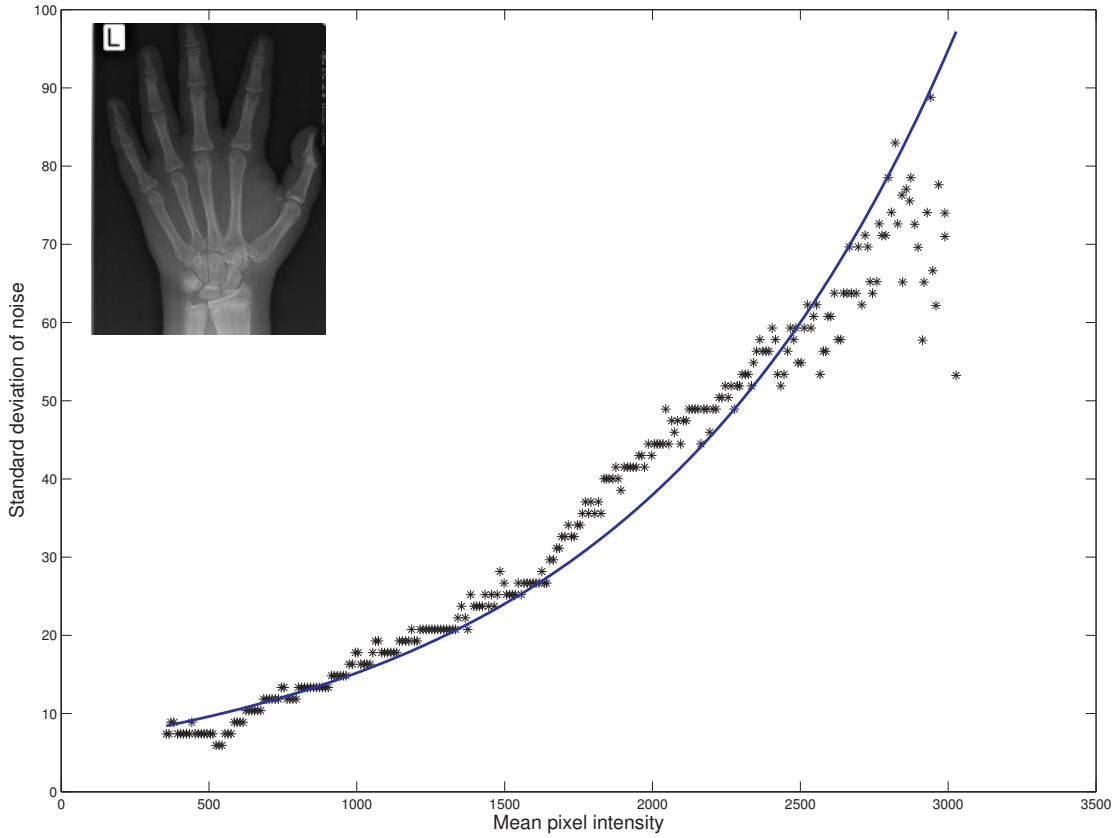


Figure 6.22 Linear regression fit (solid line) for an exponential noise model using $\log(\sigma) = a + b \cdot P$ where σ is the standard deviation of the noise and P is the mean pixel intensity. The mean pixel intensity was calculated using a 3x3 median filter for the shown radiograph. These results are for a typical hand-wrist radiograph.

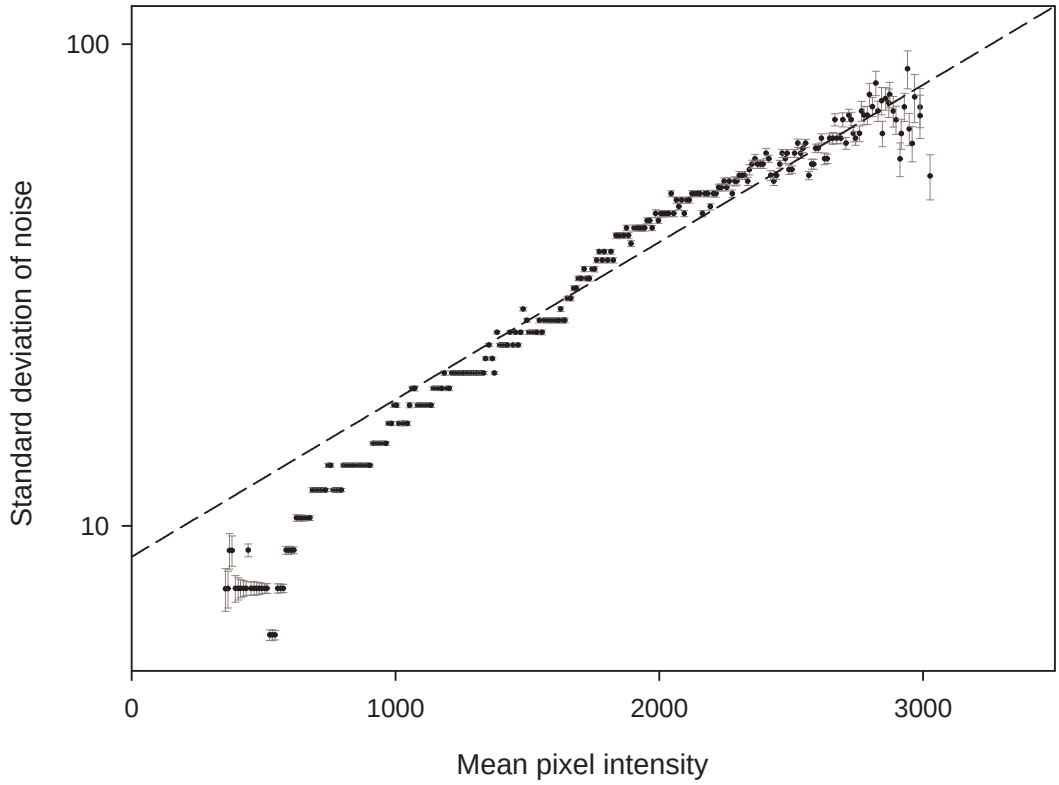


Figure 6.23 Marquardt-Levenberg nonlinear regression fitting of the noise model $\sigma = (10^a)10^{b \cdot P}$ where P is the mean pixel intensity. This is the same data as presented in Figure 6.22 except that it is on a log scale and includes error bars. The error bars correspond to one standard deviation calculated using Eqn. (6.3). The mean pixel intensity was calculated using a 3x3 median filter.

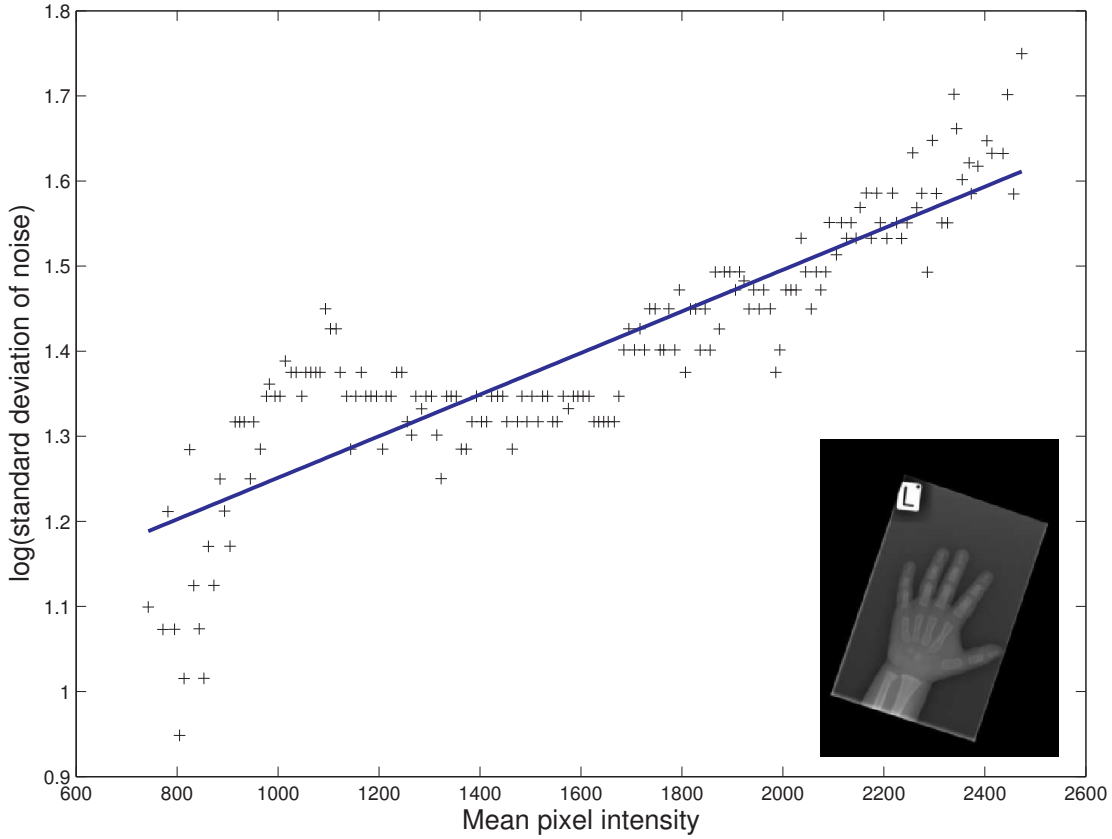


Figure 6.24 Noise estimation for a radiograph of a 30 month old child (shown). The linear regression does not model the noise variation very well, but on average there is an approximate linear relationship with pixel intensity.

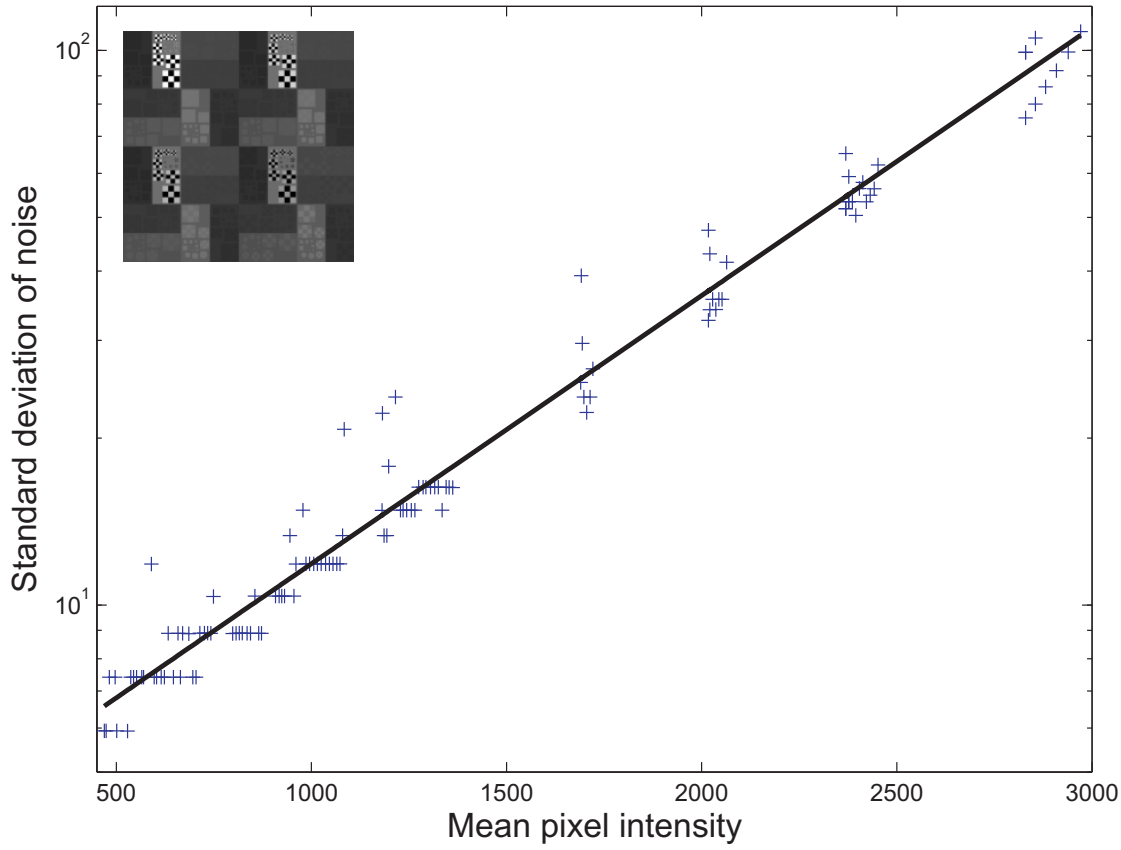


Figure 6.25 Noise estimation for the simulated TG18-BR x-ray, shown on a log scale. The linear regression line for $\log(\sigma)$ had a slope of $4.84\text{x}10^{-4}$ and intercept of 0.591. These estimated parameters compare very favourable with the theoretical of slope $5.0\text{x}10^{-4}$ and intercept 0.595 calculated from Eqn. (4.7) for $a = 1000$ and $b = 10$. Inset: The TG18-BR test pattern.

noise estimation results are illustrated in Figure 6.25. This test pattern is a good test of the noise estimation method, although it contains objects with very sharp edges and a reasonably wide distribution of pixel intensities. There is still a spread of mean-standard deviation estimates about the linear regression. This indicates that there is still opportunity for improving the estimation of the mean pixel intensities because it is these estimates that influence the binning of the pixels and, hence, the standard deviation estimates.

6.3.3 Comments on the noise estimation

The single-image noise estimation method based on linear regression gave results good enough to find approximate parameters for the exponential noise model. Furthermore, this estimation is possible irrespective of the clinical processing of the radiographs, namely the adjustment of image contrast. However, there are some significant improvements that could be made. More advanced and robust parameter estimation methods could be em-

ployed, and these are worthy of future research.

The median absolute deviation (MAD) robust estimator used for the standard deviation estimation has its limitations. From the experiments performed in this section the MAD estimator appears to have some serious limitations when handling integer distributions like the Poisson distribution. Yet for the logarithmic converted data this was much less of a problem. This limitation needs to be kept in mind when using this estimator, and it could explain some of the step effects seen on the plots of Figure 6.24 and Figure 6.25 for low pixel intensities.

The choice of a 3x3 median filter to assist in finding the mean pixel intensities was based on experiments. The largest improvement in the noise estimation could be made by finding an objective criterion for the selection of the noise estimation method. This would also help determine whether iterative methods of denoising and updating of noise models would benefit the denoising. For now, the noise model investigated in this section will be applied for signal-dependent denoising.

The linear regression parameters returned by the noise estimation procedure result in a model of the form

$$\sigma = 10^{a} \cdot 10^{b \cdot I} \tag{6.4}$$

where I is the pixel intensity, b is the slope of the line from the linear regression, and a is the intercept and acts as a scaling factor for the noise equation. This exponential noise model is used in the next section to modify the anisotropic diffusion based on the level of noise estimated for a given pixel.

6.4 Denoising using anisotropic diffusion

In this section the method of anisotropic diffusion is introduced for denoising hand-wrist radiographs. There are many methods of performing anisotropic diffusion and the method investigated in this research was based on the research by Tschumperlé *et al.* on oriented Laplacians [Tsch02]. This method was chosen because initial testing of denoising using the G'IMC extension for the GIMP software package [Tsch] showed promising results as a candidate method for denoising x-rays. A brief introduction to PDE-based image denoising is now presented, along with implementation of the oriented Laplacians method and extension of this method to include the signal-dependent noise model of Eqn. (6.4).

6.4.1 Background to the anisotropic diffusion method

Anisotropic diffusion is a nonlinear denoising method that is one member of the rapidly growing class of PDE-based methods now used for image processing. One of the advan-

tages of the anisotropic diffusion approach is the ability to adapt to image content and preserve important image discontinuities such as edges. This is achieved by using a diffusive 'flow' that is sensitive to edges and will flow along edges rather than over them, which would otherwise cause noise reduction but blurring of the edge.

The Gaussian filter is an example of a linear filter[5]. Linear Gaussian filtering is the convolution of an image $I(x, y)$ with a 2D Gaussian kernel G_σ of the form

$$G_\sigma(\mathbf{x}) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|\mathbf{x}|^2}{2\sigma^2}\right),$$

where $\mathbf{x} \in \mathbb{R}^2$ and σ is the spatial spread of the symmetric kernel. The Gaussian kernel is smooth and self-similar, meaning that it can be applied multiple times to an image and the results are predictable [Yoo04, p104]. If $I_0(x, y)$ is an initial image, then

$$I(x, y, \sigma) = I_0(x, y) * G_\sigma(x, y)$$

is the Gaussian convolution of the image. The parameter σ is sometimes referred to as the scale-space parameter. Different values of σ produce different levels of smoothing, and copies of an image at different scales is referred to as a scale-space [Sapi01, p221].

An important property of $I(x, y, \sigma)$ is that it is a solution to the parabolic linear heat flow equation, also called the Laplace equation,

$$\frac{\partial I(x, y, t)}{\partial t} = \Delta I(x, y, t) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \text{ with } I(x, y, 0) = I_0(x, y) \tag{6.5}$$

where t is an artificial time variable introduced to model the evolution of the diffusion. This solution only holds if boundary conditions are imposed on the heat flow so that there is no loss in image energy across the boundaries of the solution. This is referred to as adiabatic conditions [Yoo04, p105]. This condition is expressed as the Von Neumann boundary condition of $\partial I(x)/\partial \mathbf{n} = 0$ where $\mathbf{n}$ is the normal to the image boundary at $x \in \partial\Omega$. This solution ties together Gaussian smoothing with a PDE-based method. It can be shown that if the heat flow progresses to a time $t = T$ then the isotropic diffusion is equivalent to convolution with a Gaussian kernel $G_{\sqrt{2T}}$ [Weic08, p4] [Yoo04, p105].

Through a theoretical analysis of the method noise (that is, image - denoised image) it is possible to demonstrate that denoising using a Gaussian filter, or isotropic diffusion, results in zero noise in harmonic parts of the image. However, the noise reduction performs poorly on singular parts of the image [Buad05] [Camm09, p26], and this is an important drawback

[5]Linear filters are a collection of filters that are linear and stationary and, hence, can be implemented as a convolution with a kernel. Nonlinear filter are filters that cannot be implemented using convolutions.

of Gaussian filtering because it is the singular parts of the image, the discontinuities at edges and texture, that are often of interest. This type of image content is better handled by nonlinear filtering.

In their seminal paper Perona and Malik brought together the concepts of scale-space and a spatially varying coefficient that controls the amount of diffusion [Pero90]. They called this method anisotropic diffusion, although one has to be careful with the terminology because some authors call this nonlinear diffusion, with anisotropic diffusion having another meaning (for example, [Weic08, p2]). The isotropic diffusion Eqn. (6.5) can be written using the divergence operation ($\mathrm{div}(\cdot)$) as

$$\frac{\partial I(x,y,t)}{\partial t} = \mathrm{div}(\nabla I).$$

This specifies how the image values change with time. Perona and Malik introduced a nonlinearity into the isotropic diffusion equation through conductivity coefficients[6] $\{g : \mathbb{R} \rightarrow \mathbb{R}$ and $g(\cdot) \in [0,1]\}$ that are a decreasing function. The Perona-Malik formulation was

$$\frac{\partial I}{\partial t} = \mathrm{div}(g(|\nabla I|)\nabla I). \tag{6.6}$$

In their original paper Perona and Malika presented some recommendations on the form of the function $g(\cdot)$ [Pero90]. The function can be considered an edge-stopping function because it is usually designed to slow or stop the diffusion at boundaries. This is done by decreasing the conductivity coefficients where there are large image gradients, that is, large values of $|\nabla I|$. Figure 6.26 illustrates the behaviour of one of the functions proposed by Perona and Malik.

[6] These are sometimes referred to as diffusion coefficients or diffusivities, with corresponding diffusivity function [Weic08, p15].

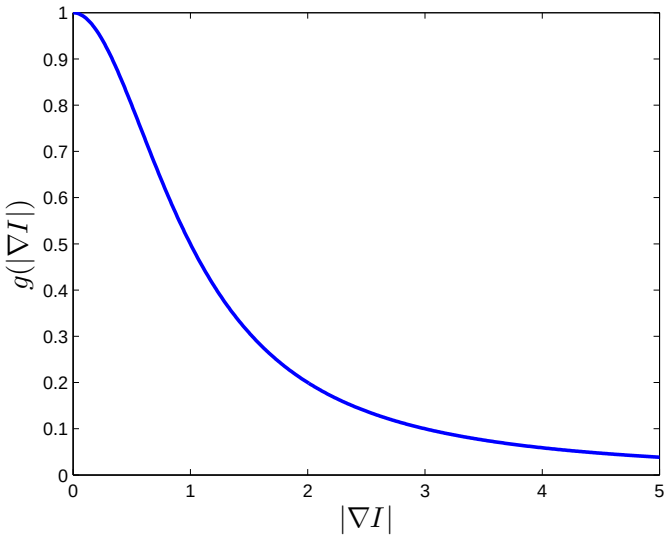


Figure 6.26 Perona-Malik conductivity function $g(|\nabla I|) = \frac{1}{1+(\frac{|\nabla I|}{K})^2}$. Small values of g result in decreased diffusion, especially across edges where $|\nabla I|$ values are large. In this example $K = 1$. The constant K is usually selected manually or set according to an estimate of the image noise [Pero90].

The Perona-Malik anisotropic diffusion has been shown to be ill-posed in the sense that two very similar starting images can appear significantly different after undergoing the same anisotropic diffusion [You96]. Interestingly, the numerical method proposed by Perona and Malik was thought to stabilise the theoretical problems with the formulation (as cited by [Tsch02, p40]). There has been a significant amount of published literature on anisotropic diffusion spawned by the Perona and Malik paper. For example, to improve the performance of the Perona-Malik anisotropic diffusion, Black *et al.* introduced the ideas of robust statistics and demonstrated the link between robust estimation methods and anisotropic diffusion [Blac98]. After initial experiments with a number of denoising methods and variants of anisotropic diffusion, the author chose to investigate and extend the oriented Laplacians method [Tsch02].

6.4.2 Anisotropic diffusion using oriented Laplacians

Anisotropic diffusion can be performed using a number of mathematical formulations. Tschumperlé showed that there are links between the three formulations of regularising functional minimisation, divergence-based diffusion, and oriented Laplacian diffusion, that is, diffusion using using directional second derivatives [Tsch05] [Tsch02].

Central to the formulation of oriented Laplacians is the idea that anisotropic diffusion of

the image I can be decomposed into two orthogonal components[7] I_{uu} and I_{vv} where the the second derivatives of I are in the directions $\mathbf{u}$ and $\mathbf{v}$ ($\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ and $\mathbf{u} \perp \mathbf{v}$) and

$$I_{uu} = \mathbf{u}^T \mathbf{H} \mathbf{u} \quad \text{and } I_{vv} = \mathbf{v}^T \mathbf{H} \mathbf{v}$$

where $\mathbf{H}$ is the Hessian of I given by $\mathbf{H}_{i,j} = \frac{\partial^2 I}{\partial x_i \partial y_j}$

This decomposition leads to the oriented Laplacians equation

$$\frac{\partial I}{\partial t} = g_1 I_{uu} + g_2 I_{vv} \tag{6.7}$$

where $g_1, g_2 > 0$.

This decomposition allows the anisotropic diffusion to be thought of as the combination of two oriented heat flows or two oriented diffusion Laplacians with different conductivity coefficients (smoothing weights) in the directions (g_2) of the gradient, and orthogonal to the gradient (g_1). The g_1 and g_2 are usually functions of the image gradient and define a large range of anisotropic diffusion methods. For example, using $g_1 = 1, g_2 = 0, \mathbf{v} = \frac{\nabla I}{|\nabla I|}$ and $\mathbf{u} = \mathbf{v}^\perp$ the oriented Laplacian equation Eqn. (6.7) becomes

$$\frac{\partial I}{\partial t} = I_{uu}$$

This diffusion along the lines orthogonal to the gradient is called the mean curvature flow [Weic08, p39].

One of the advantages of adopting the oriented Laplacians formulation is that they have a geometric meaning that make it easier to understand the nature of the smoothing [Tsch05]. The other major advantage is that it can be shown that

$$\frac{\partial I}{\partial t} = g_1 I_{uu} + g_2 I_{vv} = \mathrm{trace}(\mathbf{TH}) \tag{6.8}$$

where $\mathbf{T} = g_1 \mathbf{u}\mathbf{u}^T + g_2 \mathbf{v}\mathbf{v}^T$. This is referred to as the trace-based form for the anisotropic diffusion.

The 2x2 symmetric matrix $\mathbf{T}$ is called a diffusion tensor and has eigenvalues g_1 and g_2 for respective eigenvectors $\mathbf{u}$ and $\mathbf{v}$. Diffusion tensors are a structure tensor in which the eigenvalues of the structure tensor have been used to regularise the diffusion, that is, change the local behaviour of the anisotropic diffusion so that it is not ill-posed.

The role of the diffusion tensor can be seen by first starting with the original Perona–Malik

[7] The terminology adopted here is similar to that used by Tschumperlé [Tsch02]

formulation. The Perona–Malik form of the diffusion equation uses an edge-stopping function based on a scalar quantity. In this regards it uses a nonlinear function, but is isotropic (the same behaviour in every direction for a given gradient $|\nabla I|$). The diffusion tensor, on the other hand, allows the local orientation of the gradients to influence diffusion because of the link with the structure tensor.

The (isotropic) nonlinear diffusion uses a gradient magnitude based on $\nabla I^T \nabla I$. Although it is possible to work with gradients directly, this is usually not practical because if one wants to regularise or smooth gradient values to make them useful, cancellation of gradients can occur. Brox *et al.* present an example of a thin line with a positive gradient on one side and a negative gradient on the other [Brox06]. If a smoothing operation is performed, then the two gradients cancel each other out. The way to get around this is to form a tensor $\mathbf{S}$ (in this case a matrix) from the outer product $\nabla I \nabla I^T$ [Brox06] [Weic08, p56]. This tensor can then be smoothed without the cancellation effects.

Performing an eigen decomposition of the tensor $\mathbf{S} = \nabla I \nabla I^T$ gives eigenvalues $\lambda_1 = |\nabla I|^2$ and $\lambda_2 = 0$. In this situation little has been gained from using the outer product. The advantage arises when smoothing is applied to the tensor. Gaussian smoothing G_σ of a structure tensor results in a tensor $\mathbf{S}_\sigma = \mathbf{S} * G_\sigma$ where the value of the tensor at point (x, y) is influenced by the neighbourhood of (x, y). This improves the orientation information that can be derived from the structure tensor [Brox06].

The smoothed structure tensor $\mathbf{S}$ can be used to provide direction-specific smoothing for the diffusion by using the eigenvalues of the tensor in the edge-stopping function $g()$. This gives the so-called diffusion tensor [Weic08]. Directions in which there are large variations in local structure have a correspondingly large eigenvalue λ^+ in the structure tensor. When these large values are used in the edge-stopping function, $g(\lambda^+)$ is small, so diffusion in the direction of the corresponding eigenvector θ^+ slows or stops. This stops the diffusion at discontinuities and other locations with large gradients. Orthogonal to this direction the eigenvector θ^- with eigenvalue λ^- will be small, so $g(\lambda^-)$ is large and diffusion will occur in the direction of θ^-. This results in little blurring across edges, and a dominance of blurring along edges. This results in anisotropic diffusion that is based on orientation of the edges.

Critical to the success of the anisotropic diffusion is the selection of an edge-stopping function. This is now a function of the diffusion tensor eigenvalues λ_i rather than the magnitude of the image gradients $|\nabla I_i|$. Eqn. (6.8) can be written [Tsch06a]

$$\begin{aligned} \frac{\partial I}{\partial t} &= \operatorname{trace}(\mathbf{TH}) \\ &= \operatorname{trace}\left((g^-_{(\lambda^+,\lambda^-)} \theta^- {\theta^-}^T + g^+_{(\lambda^+,\lambda^-)} \theta^+ {\theta^+}^T)\, \mathbf{H} \right) \end{aligned} \tag{6.9}$$

where the eigenvectors and eigenvalues are pointwise within the image.

Although theoretical developments are occurring for the selection of an optimal edge-stopping function for a given application, it still remains mostly a matter of trying different functions for the application. The function used in this research was based on the published successful use of the function [Tsch06a], including being used for the detection of wire structures in x-rays [Tsch06b]. The function is of the form

$$g^{+}_{(\lambda^{+},\lambda^{-})} = \frac{1}{(1+\lambda^{+}+\lambda^{-})^{p2}} \quad \text{and} \quad g^{-}_{(\lambda^{+},\lambda^{-})} = \frac{1}{(1+\lambda^{+}+\lambda^{-})^{p1}} \quad \text{with } p1 < p2. \tag{6.10}$$

This function allows for control of the diffusion in both the direction of maximum variation and the direction of minimum variation, with the additional couple of the two terms via the eigenvalues.

6.4.2.1 Incorporating a signal-dependent noise model into the anisotropic diffusion

The signal-dependent noise model developed in Section 4.3.1 showed that higher pixel intensities in the image have a larger associated noise. The parameters for this model were found for each radiograph using the method described in Section 6.3. The resulting standard deviation of the noise is of the form $10^{b \cdot I(x,y)}$ where $I(x,y)$ is the pixel intensity at point (x,y) in the image. Given the estimate of b from the noise estimation method, it is possible to estimate the standard deviation of the noise. The point in question is what to do with this estimate.

Ideally, the parameters $p1$ and $p2$ should be set according to the noise content of the image. However, the author found that these parameters had a significant effect upon the quality of the anisotropic diffusion - the results were too sensitive to the values chosen. It is likely than another type of diffusivity function will be required to reduce this sensitivity, and this has been reserved for future research. In the meantime, to incorporate the signal-dependent noise model into the anisotropic diffusion method, a new noise slowing function $s(I,t)$: $s \in [0,1]$ was incorporated into Eqn. (6.9)

$$\frac{\partial I}{\partial t} = s(\sigma,t)\,\mathrm{trace}\left((g^{-}_{(\lambda^{+},\lambda^{-})}\theta^{-}\theta^{-^{T}} + g^{+}_{(\lambda^{+},\lambda^{-})}\theta^{+}\theta^{+^{T}})\,\mathbf{H}\right). \tag{6.11}$$

As with the stopping function $g()$ for the Perona–Malik anisotropic diffusion, the intention was to slow the diffusion in areas of low noise so that the diffusion time t could be extended and the diffusion would favour noise reduction in high noise areas. The form of such a function is not obvious. Furthermore, the noise model of Eqn. (6.4) only applies to the radiographic image before denoising. This means that once denoising commences the model of the relationship between noise and pixel intensity will no longer apply. Although

it should theoretically be possible to track the influence of the denoising on local statistics, there is currently no theoretical development to support this. As an alternative, the approach was to have the function $s()$ depend on the diffusion time in such a way that by time T, $s(I,T) = 1$, corresponding to the normal diffusion equation. The chosen function is illustrated in Figure 6.27 and is of the form

$$s(I,t) = \frac{t}{T} + \frac{10^{b \cdot I}}{10^{b \cdot I_{max}}}(1 - \frac{t}{T}) \tag{6.12}$$

where T is the total diffusion time (total number of iterations for the denoising), $I \in [0, 4095]$, and $I_{max} = 4095$, the maximum pixel intensity for the DICOM images.

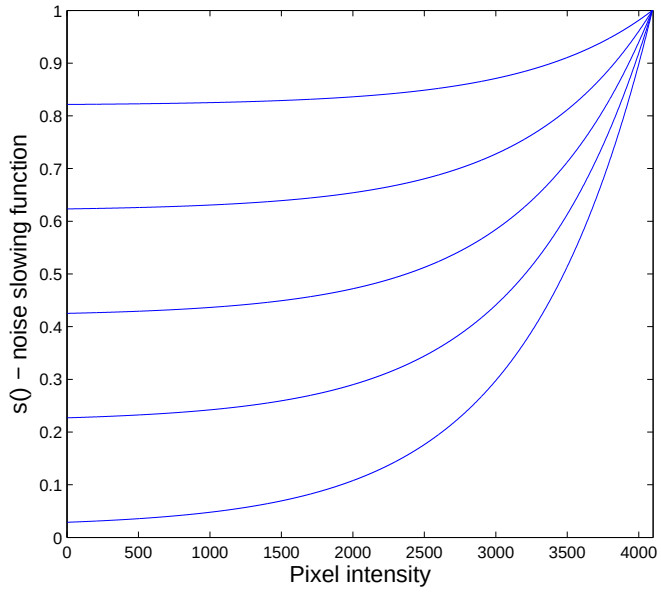


Figure 6.27 The noise slowing function for the anisotropic diffusion. This is an example of 50 iterations with the lines shown every 10 iterations. The function is for the theoretical noise of b = 1/2000

6.4.3 Implementation of the oriented Laplacians anisotropic diffusion

The oriented Laplacians method was implemented using the CImg software package [Tsch] which the author linked with MATLAB® to provide some fast image processing tools. The numerical details of the implementation have been previously described by Tschumperlé [Tsch02, p101].

The regularisation control parameters of $p1 = 0.5$ and $p2 = 0.9$ were found empirically for the hand-wrist radiographs. Figure 6.28 shows an example of anisotropic diffusion filtering of a carpus image using these parameters without the signal-dependent term. The param-

eters $p1$ and $p2$ required careful fine tuning, as did the structure tensor smoothing. The quality of the anisotropic diffusion was very sensitive to these parameters. This obviously requires attention in future research and consideration of alternative robust estimators for anisotropic diffusion. For the rest of the results in this thesis, unless stated otherwise the images were denoised using the oriented Laplacians anisotropic diffusion method for 200 iterations ($T = 200$), with Gaussian smoothing of the gradient using $\sigma = 1.0$ prior to calculating the structure tensor and again afterwards with $\sigma = 0.7$. The only smoothing of the actual image data was that due to the anisotropic diffusion process. The performance of the oriented Laplacians is evaluated in the next section.

(a) Original image (outlier filtered)

(b) Image denoised using oriented Laplacians

Figure 6.28 Anisotropic diffusion filtering of a carpal image using parameters $p1 = 0.5$ and $p2 = 0.9$ and 50 iterations. (a) original image after outlier filtering, (b) anisotropic diffusion filtered image. Note the significant reduction in noise in the image, but with preservation of edges.

6.5 Measuring the performance of denoising methods

There is no standard performance measure for image denoising methods. This partly relates to the different objectives of the denoising methods and to the significant differences in structure and noise content of images. The performance measurement is usually broken into the tasks of assessing the amount of noise reduction and assessing the impact the denoising has on image blur and artefact generation. The common approach to measuring the noise reduction is to estimate the noise in the original and processed images and to measure the difference between the two using a metric such as the mean-square error. However, such a measure does not always correlate with the human visual assessment of the noise. Displaying the difference between the original image and the denoised image (the so-called 'method noise') is useful for visually assessing the artefacts in the smoothed image. It is also useful for checking that the noise component is as similar to white noise as possible and does not contain structural components [Buad05] . What this method does

not assess is the preservation of edges that are important for edge-based image segmentation and feature extraction. One possible approach to assessing the preservation of edges is to use a sharpness metric that is insensitive to image noise. Thirteen sharpness metrics have been reviewed by Ferzli and Karma [Ferz09] and these include methods such as derivative-based metrics, histogram entropy-based metrics, and frequency-based metrics. Most of these metrics are directed at the problem of autofocussing systems where the degree of blur is used to control the focusing of some system. In such systems the blur is often modelled as a convolution with a single point spread function that is spatially stationary and so it is only necessary to estimate this single spatial spread across the whole image. This is not the problem that is being addressed in the application of a sharpness metric to denoising, and so many of the sharpness metrics are unsuitable for quantifying the blur caused by denoising algorithms.

A number of initial experiments were performed using candidate methods for measuring image blur in the presence of noise. Although there are some possible contenders for this application, such as [Wee08] and [Zhu09], the results were not very successful and further research is required in this area to find a reliable and general metric. The author decided to visually determine the impact of the denoising on image blur both from inspection of bone edges and from changes to the edges detected using the Canny edge detector [Cann86] from the MATLAB® Image Processing Toolbox (default parameters: $\sigma = 1$ and automatic finding of threshold based on a 70% criteria for the percentage of non-edge pixels – this is reasonable for hand wrist radiographs). The mean Structural Similarity Index (SSIM) [Wang04] was used to quantify the amount of noise reduction by comparing the original with the denoised image. The image energy, $E = \sum_{\Omega} I(x,y)^2$, was checked to make sure the images did not differ so much in pixel intensity that the SSIM measurement was unreliable. As with the measurements in Section 6.2.2.2, the SSIM measurement used the default parameters recommended by the author [Wang04].

6.5.1 The performance of anisotropic diffusion using a model of signal-dependent noise

The first comparison of performance was the Canny edge detection applied to the original and denoised radiograph. As illustrated in Figure 6.29 the denoising method significantly improves the performance of the Canny edge detection. Increasing the σ for the Canny edge detector does improve the detection of edges, but this has the typical drawback of an isotropic operation in that there can be misplacement of edges.

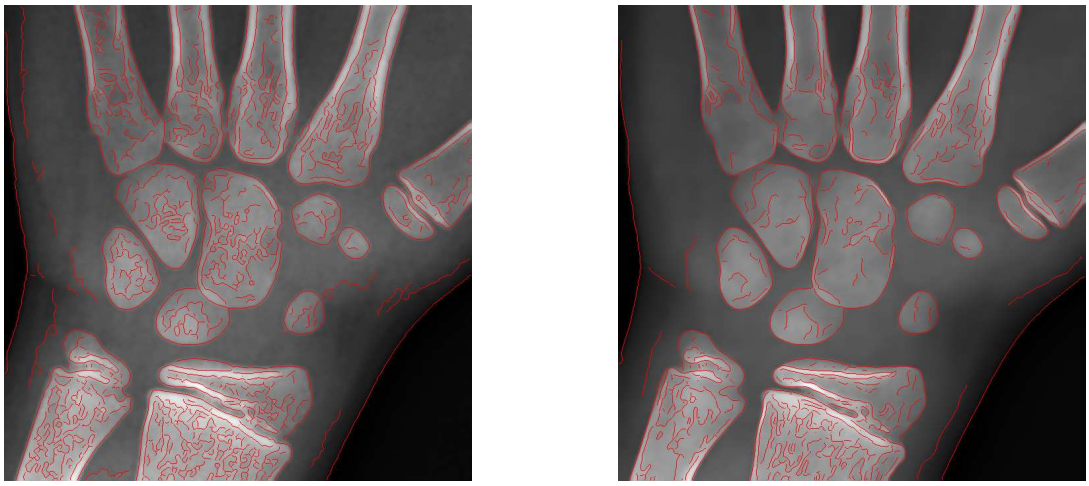
The denoising filter was also compared with an existing Perona–Malik method of anisotropic diffusion denoising that has been applied to the carpal bones of young children prior to Canny edge detection [Zhan07], and applied to the carpal bones prior to segmenting using

(a) Original radiograph (b) Denoised radiograph

Figure 6.29 Comparison of of the Canny edge detection results for (a) original radiograph, and (b) radiograph denoised using the oriented Laplacians anisotropic diffusion method ($\sigma = 1$ for the Canny edge detector).

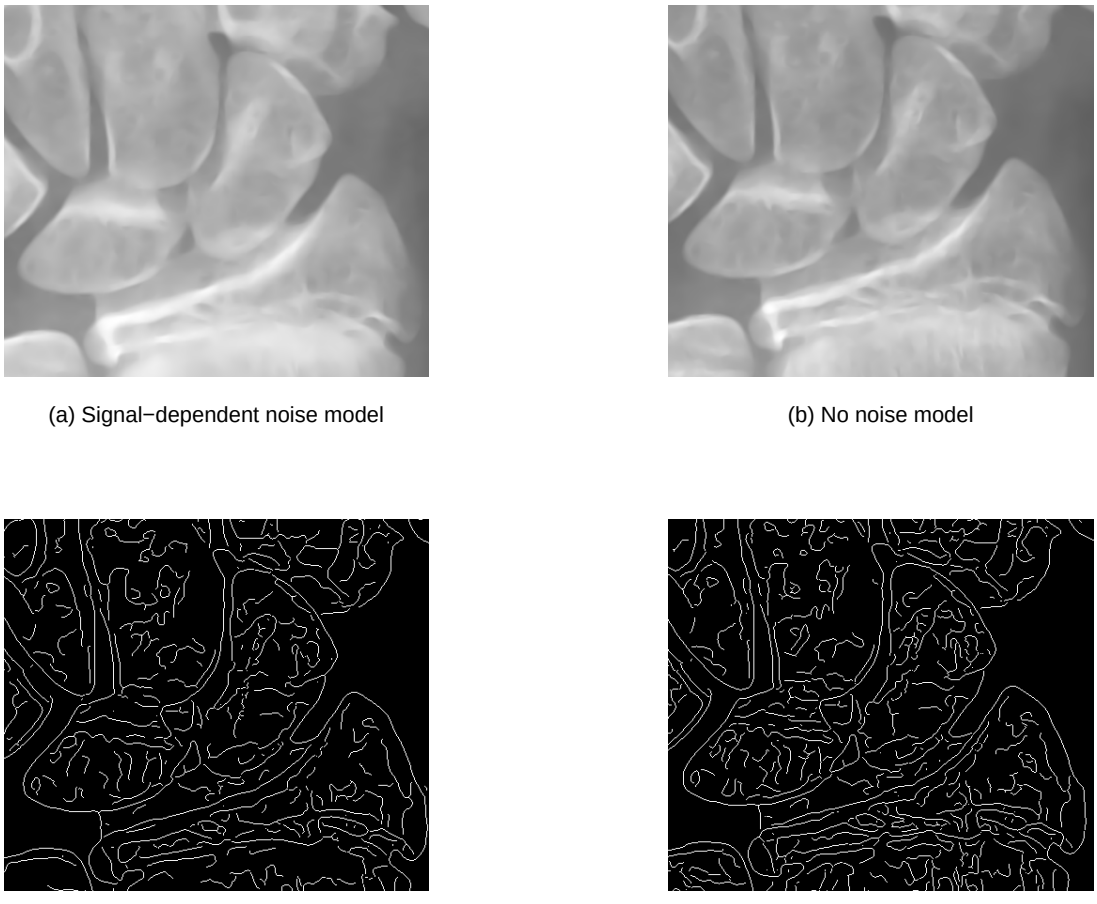
the level set method [Lin05]. This method was applied to the bones of an older child and the compared with the oriented Laplacians method. Figure 6.30 illustrates the superior performance of the oriented Laplacian method. The bone edges are sharper and the interior noise of the bones significantly reduced.

The performance of the oriented Laplacians method was tested with and without the noise model by applying the filter to the carpal region from all eight images from the development dataset (see Appendix E). The edges detected using the Canny edge detector were displayed alongside the original image (see Figure 6.31) and the edge detection performance classified according to improvement (+), no change (0), or diminished performance (-) based on the continuity of the contours and spurious contours detected. The results for the eight radiographs from the development dataset are listed in Table 6.2. Recall that larger values of MSSIM indicate that the image is more like the original. In this case it means that there has been less noise reduction, or if the blur is significant, the value of MSSIM will decrease (as illustrated in Figure 6.7). What is clear from the results is that both the denoising methods significantly decreased the noise, although this was also obvious qualitatively. What was also clear from the results was although the signal-dependent denoising did preferentially denoise the bones slightly more, meaning it was operating correctly (see Figure 6.31), the effect of the signal-dependent term was not very significant when it came to edge detection using the Canny edge detector.



(a) (b)

Figure 6.30 Comparison of anisotropic diffusion filtering using the Perona–Malik formulation with the oriented Laplacians method. The indicated contours are the edges found using the Canny edge detector ($\sigma = 1$) (a) The method used by Zhang *et al.* [Zhan07] and Lin *et al.* [Lin05] for denoising the carpal bones prior to segmentation; (b) the oriented Laplacians method has sharper bone contours and less noise in the interior of the bone.



(a) Signal−dependent noise model (b) No noise model

(c) Canny edge detection (d) Canny edge detection

Figure 6.31 The method of comparing the 'signal-dependent noise model' and 'no noise model' oriented Laplacians denoising methods.

Table 6.2 Performance comparison for the 'signal-dependent noise model' and 'no noise model' oriented Laplacians methods for image denoising. The MSSIM reflects the amount of noise reduction compared with the original radiograph. The visual comparison of the methods was a grading of the Canny edge detection results according to improvement (+), no change (0), or diminished performance (-) of the anisotropic diffusion for the signal-dependent noise model (SDNM) compared with the no noise model (NNM).

Patient	MSSIM for SDNM	MSSIM for NNM	Edge detection (visual)
1	0.912	0.916	0
2	0.828	0.861	+
3	0.903	0.927	-
4	0.854	0.858	0
5	0.929	0.927	0
6	0.894	0.907	+
7	0.886	0.881	0
8	0.958	0.958	+

6.6 Conclusions

Impulse noise is a nuisance in image processing and a method was developed in this chapter for removing this noise through a process of detection and correction. The median absolute deviation (MAD) robust statistic was used to identify outliers for removal. Through simulation it was demonstrated that the MAD-based outlier filter works very well compared with both the standard deviation-based outlier filter and the star-shaped median filtering.

Utilising the noise model developed in Chapter 4, it was possible to estimate parameters for the radiographic noise by binning pixel intensities and using the MAD estimator to find the standard deviation in each bin. The performance of the MAD estimator was analysed extensively in this chapter, especially because the computed radiography (CR) system performs a logarithmic conversion of the pixel intensities and this changes the statistical distribution of the data. The MAD estimator performed well although it showed an unusual 'step' characteristic for Poisson noise that makes it worthy of further investigation. This was less important for noise estimation for the radiographic images because of the logarithmic conversion of the pixel data.

Although the noise estimation method requires a more robust approach for parameter estimation, the results showed that the method worked remarkably well for such a relatively simple implementation compared with methods that eliminate edges from the analysis. This allowed noise to be estimated on a pixel-by-pixel basis using the simple but robust median filter for finding the mean pixel intensity for a small sampling neighbourhood. This

noise estimate was used in a new approach to anisotropic diffusion filtering in which the level of noise in a pixel controls the amount of local diffusion. The resulting filter was sensitive to the parameters selected for the filter and the results for denoising of the hand-wrist radiographs only showed a marginal improvement in the image noise level and results for edge detection. However, the implementation of the signal-dependent noise control in the diffusion process was rather simplistic and investigation of alternative methods of control is recommended. Even without the signal-dependent noise term, the oriented Laplacians anisotropic diffusion was very good at denoising the radiographs and preservation of edges, and performed much better than the Perona–Malik method that has previously been used for denoising images prior to carpal bone segmentation [Zhan07] [Lin05].

The Mean Structural Similarity (MSSIM) index was used to measure the performance of both the impulse noise removal method and denoising of the carpal images. This index decreased with the amount of added impulse noise and increased with impulse noise removed by a MAD-based outlier filter. The index has limitations when applied to impulse noise removal, the main one being that it is sensitive to incomplete impulse noise removal. For example, when 5x5 pixel impulses were filtered using the MAD-based outlier filter with a 5x5 pixel neighbourhood there was incomplete removal of the impulses, yet the MSSIM value still increased slightly. When the MSSIM index was used to measure the performance of image denoising with a signal-dependent term in the anisotropic diffusion equations, there was a slight decrease in the MSSIM index indicating improved denoising. In this application the MSSIM index decreases with the amount of denoising. However, the MSSIM index also decreases with image blur so the slight decrease could have been due to blur caused by the anisotropic diffusion rather than improved denoising. This makes the MSSIM index undesirable for assessing denoising when blurring is possible. What is required is an index that measures both denoising and preservation of edges simultaneously, or a measure of denoising and a measure of edge blur that are independent of each other and can be combined as a single metric. Because there is currently no such metric available, in future work it will be necessary to optimise the denoising method through the standard approach of determining the impact on a image analysis task, in this case the success of segmentation of the carpal bones.

Chapter 7

Conclusions and future work

This chapter is a summary of the main contributions of the thesis and conclusions that can be drawn from the research that was undertaken. Detailed discussion is contained in the relevant chapters. The conclusions are followed by suggestions regarding future research.

7.1 Conclusions

For bone age assessment to be of use in both clinical and research environments, it is important that the bone age estimate be accurate and reproducible between assessors (observers). There are three main methods of bone age assessment: Greulich and Pyle; Tanner and Whitehouse; and the Fels method. From a review of these three methods, there is no conclusive evidence that any one method is better than another: all three methods suffer from relatively large intra- and inter-observer variability, and the accuracy of each decreases outside of the reference populations from which they were developed. Furthermore, bone age assessment is a relatively complex and time consuming process and there have been numerous approaches to computerise assessment, with the main aim being to improve the accuracy and reproducibility of the assessment. With an accurate and reproducible automated method it would be much easier to performing large-scale studies to address the reference population-specific concerns about bone age assessment, and to clarify the role of bone age assessment in the routine monitoring of therapeutic intervention, especially its role in children receiving growth hormone.

The different approaches to computerised assessment of bone age have ranged from computer-assisted manual methods, through to fully automated assessment. Most of these methods have implemented aspects of the Greulich-Pyle or Tanner-Whitehouse bone age assessment methods (no published research on automated implementation of the Fels method was found). The area of computerised bone age assessment that has received the least at-

tention is the carpal bones. From a review of the literature, an argument was presented for investigating the role of the carpal bones in computerised assessment of bone age, especially the role of the carpal bones in older children. It was argued that even if the sparse amount of existing evidence for excluding the carpal bones in manual methods is considered reasonable, this evidence may not be applicable to the computerised assessment of bone age. The potential benefits from computerised analysis are sufficient to warrant investigating the role of the carpal bones in bone age assessment, especially bone age assessment in children in the difficult period over nine years of age when these bones overlap.

One especially challenging aspect of analysis of the carpal bones is segmentation of the bones - finding of the bone contours. In addition to carpal bone overlap and poor contrast of these bones against the thickness of soft tissue of the hand, the radiographs usually contain a significant amount of noise. Reducing the influence of this noise on image processing tasks is necessary. To this end a signal-dependent noise model was developed, noise parameters found for this model, and the model used in a method of image denoising based on anisotropic diffusion.

The first step in developing the signal-dependent radiographic noise model was to understand the pixel intensity and noise characteristics of the image receptor. This was performed by designing and constructing a hand phantom that could in used to investigate the image characteristics of hand-wrist radiography using the Kodak CR 900 system. The computed radiography system was found to be very good at maintaining pixel intensities across a range of x-ray doses to the computed radiography plate, although at very low exposures the system implemented a low-dose optimisation algorithm that significantly influenced the pixel intensity and noise characteristics. This constancy of pixel intensities was used to identify artefacts in the images, such as radiographic markers. Even with this constancy of pixel intensities, an analysis pixel intensity histograms for 97 hand-wrist radiographs showed that was not possible to separate bone from soft tissue based on pixel intensity differences alone.

Important for the task of segmenting overlapping carpal bones was the finding that the bone overlap is not a simple addition of pixel intensities, and that the contrast of the overlap region changes with the bone thickness, that is, pixel intensity. This can be explained by the physical process of energy components of the x-ray beam being selectively attenuated as the x-rays pass through thicker sections of bone - the x-ray beam is said to 'harden' because 'soft' (low energy) x-rays are removed. This means it is not possible to use a simple subtraction of pixel intensities to simplify the segmentation of the overlapping bones.

The noise characteristics of the computed radiography system were modelled using a simple, single Poisson noise model, but with compensation for the logarithmic conversion that

happens before the pixel data is present at the output of the computed radiography system. The derived theoretical model was an exponential relationship between the variance of the image noise and the pixel intensity. This theoretical model was validated using Monte Carlo simulation, radiographs of a stepwedge, and homogeneous regions within the radiographs of the hand phantom. Although this was for the Kodak CR 900 system, many of the other manufacturers use a similar logarithmic conversion, so with suitable validation, the model should be transferable to these other systems.

In order to use the exponential noise model it was necessary to estimate noise parameters for individual images. However, it was first necessary to remove the impulse noise that is a nuisance in image processing. The method used in this research was an outlier filter based on the process of detection and correction. The impulses in the computed radiography images were clusters of 1-10 connected pixels, with the most common being three pixels. The impulse noise was detected as a statistical outlier using the robust statistic of median absolute deviation (MAD) for estimating the standard deviation of the pixels in a neighbourhood. Outlier pixels were then replaced with the median of the neighbourhood because the median is a more robust statistic than the mean. For such impulses the median absolute deviation (MAD) based outlier filter out-performed common impulse noise filters of median filtering, star-shaped median filtering, and standard deviation-based outlier filtering. This superior performance was in terms of impulse noise removal and minimal impact on the image noise distribution. The minimal impact on the noise distribution was required so there was little effect on the subsequent noise parameter estimation for the image.

With the impulse noise removed, the image denoising was performed in two stages: estimation of the noise parameters, and application of the method of anisotropic diffusion using a signal-dependent noise model. The signal-dependent noise model was applied because it is known that although anisotropic diffusion aims to reduce the blurring of image edges, this still occurs, especially where the edge contrast is initially low.

The noise parameters for each image were estimated using a single-image noise estimation method based on a collection of mean-standard deviation pairs. This used a new approach of binning pixel intensities for every pixel in the hand soft tissue contour, rather than avoiding areas where there were structures that could upset the estimation of the standard deviation. By only having to estimate the mean pixel intensity to bin the pixels, the noise measurement was much less sensitive to image structure than, say, the quadratic nature of a standard deviation estimate using small spatial neighbourhoods. The MAD estimator was used as a robust statistic for estimating the variance of the noise in the pixel intensity bins. The performance of the MAD estimator was extensively tested for the logarithmic pixel intensity characteristics of the computed radiography images. With the exception of

very low exposures, Monte Carlo simulations verified that the majority of the noise distribution in the computed radiography images could be modelled as a signal-dependent Gaussian distribution with an estimate of the noise found using the MAD estimator. The noise estimation method performed well for both synthetic and clinical images, although the results were poorer for images of very young children where there was very little bone area in the carpus, and most of the carpal bones were not yet ossified.

Having developed a method to estimate the parameters of the signal-dependent noise model for an image, a denoising method was developed that used this signal-dependent model. This method was based upon diffusion using oriented Laplacians where the diffusion across edges was penalised and diffusion along edges favoured so that less edge blurring occurred. Even with this approach it was observed that at low contrast edges the diffusion would bleed through the bone contour and the edge was lost. A signal-dependent term was introduced whereby the diffusion was slowed based on the estimated noise component. The idea was to allow more diffusion to take place in areas that required it (high noise), and less in areas that did not require it (low noise). Measuring the performance of this denoising method was difficult without a reference image. The usual approach to assessing the performance is to examine the impact of denoising on the performance of the segmentation method. The approach taken in this research was to investigate decoupling this dependence on the segmentation method and to investigate methods of optimising the denoising in isolation so that the performance of the segmentation could be optimised separately. A number of different methods were investigated but no suitable reference-free image metric was found for assessing both denoising and the impact on blur. In the absence of such a metric the results of a Canny edge detector were assessed qualitatively as a measure of edge blur and the reduction in noise in the image measured using the Mean Structural Similarity (MSSIM) index. It was found that adding the signal-dependent term to the anisotropic diffusion resulted in slightly more diffusion in areas of high noise (namely, the bones) producing a slight reduction in noise in some of the radiographs compared with having no signal-dependent term, although the improvement was only marginal. Even without the signal-dependent noise term the oriented Laplacians anisotropic diffusion was very good at denoising the radiographs and preserving edges, and it performed much better than the Perona–Malik method that had previously been used for denoising images prior to carpal bone segmentation.

7.2 Future work

At very large x-ray exposures the expected amount of radiographic noise drops below that predicted by the exponential noise model. This can be explained by a departure of the computed radiography pixel intensities from a simple logarithmic relationship: the response

'rolls-off' with increasing dose to the computed radiography plate. A more complex pixel intensity and noise model may be required to overcome this limitation, although it should be kept in mind that x-ray exposures at which this starts to happen are not used clinically for hand-wrist radiography.

The hand-wrist radiographs are processed in the computed radiography system using algorithms that require further investigation for their impact on image noise. The difficulty is that each of the computed radiography systems from the different manufacturers apply different image processing algorithms, so what is required is a general model of image noise that can be adapted to the individual computed radiography systems. It is likely that this adaptation would need to be applied based on the model of computed radiography system (from the DICOM file header information). It is unlikely that it will be possible to estimate the parameters of the model from a single clinical radiograph, unless the model can somehow be generalised across a range of computed radiograph systems.

The original aspect of the research on impulse noise removal was the simplicity and robustness of the method in comparison with more elaborate methods of correcting impulse noise in the presence of other noise sources, such as the method used by Frosio and Borghese for the correction of impulse noise in the presence of Poisson noise [Fros09]. The drawback with the simplistic approach was that the impulse removal method resulted in detection and correction of more random impulses in areas of the image with little or no edge or texture components than would be expected for a Gaussian distribution of pixel intensities. These false detections could be reduced by increasing the size of the neighbourhood used for the calculation of the median absolute deviation. However, an increased neighbourhood size is undesirable because the pixel intensity calculation for the correction of outliers is no longer localised. An alternative approach may be to expand the neighbourhood size, but to spatially weight the pixel intensities in the neighbourhood to improve spatial localisation for both detection and correction. This is similar to the idea used for non-local means image denoising [Buad08] where each pixel is given a weight that decreases with increasing distance from the centre of the neighbourhood being considered.

A non-weighted linear regression was employed for finding the parameters for the single-image noise estimation. If background pixels are accidentally included in the region of interest used for the noise assessment they can significantly change the linear regression fit. This limitation may be overcome by using a nonlinear curve fitting method with weighted regression based on confidence of the bin estimates, so that small background regions will have little impact on the fit. Furthermore, the results for the synthetic radiograph based on the AAPM TG18-BR test pattern showed that there is still a spread of mean-standard deviation estimates about the line of linear regression. This indicates that there is still opportunity to improve the estimate of the mean pixel intensities because it is these esti-

mates that influence the binning of the pixels and hence the standard deviation estimates. It is also possible that an iterative successive refinement of the mean pixel intensities may improve the noise estimation. This would involve estimating the noise, denoising of the image using this estimate, then once again eatimating the noise using the denoised pixel intensities to bin the original image data.

Part of the spread in the mean-standard deviation estimate for the TG18-BR test pattern image could also be due to the MAD estimate itself. The MAD estimator showed a step behaviour when simulations were performed using a discrete Poisson random number generator. Although these steps were less pronounced when logarithmic conversion of the data was performed (to simulate the computed radiography system), this could be a significant limitation of this estimator and further investigation is required.

It was observed that the performance of the signal-dependent anisotropic diffusion method was very sensitive to the selection of parameters. It is likely that the cause of this problem is the function that has been used to incorporate the noise model information into the diffusion equation. A new noise function is required and this should include analysis of the stability of the resulting partial differential equation scheme employed in the oriented Laplacians diffusion. Furthermore, the 'stopping function' implementation of the signal-dependent noise control in the diffusion process was rather simplistic and investigation of alternative methods of control is recommended, including using the noise function to control the overall parameters of the diffusion rather than diffusion at a pixel level. This may improve the stability of the anisotropic diffusion process.

Optimising the denoising method usually involves applying the method then applying the segmentation method and determining the influence of the denoising on the segmentation results. Unfortunately, this means that the results can be significantly influenced by the segmentation method, so there is no truly independent assessment of the denoising method. There is currently no reference-free metric for measuring the quality of denoising when anisotropic diffusion is performed. This is worthy of further research because such a metric would greatly assist the development and optimisation of the denoising method.

Appendix A

A review of the performance of the computer-aided skeletal age scoring system (CASAS)

This appendix presents a summary of the scientific literature comparing performance of the computer-aided skeletal age scoring system (CASAS) with the manual Tanner-Whitehouse (TW) bone age method. Most of the studies looked at either TW bone stage differences or calculated bone age differences between the two methods. Although to a limited degree it is possible to compare the bone stage differences with the calculated bone age differences, what is most surprising is that only a few studies have performed a comparison between the manual TW radius-ulna-short bone (RUS) method and the CASAS system. The most extensive comparison appears to have been that performed by van Teunenbroek *et al.*, but their study presents a mix of CASAS bone age results using both the 13-bone TW RUS method, and the less time-consuming 6-bone method [Teun96]. Even so, Table A.1 shows that there is reasonable agreement between the manual TW and CASAS results for both healthy children and children with some growth disorders. However, the mean difference is only one comparison and it provides no measure of the spread of the bone age values. With the exception of studies by Tanner *et al.*, Table A.2 shows that the CASAS system can produce differences of up to two stages for repeat assessments, but the no- and single- stage differences are comparable with those for the manual method.

The number of bone assessments requiring manual insertion of a bone stage averaged 9.3% across different studies and pathologies (Table A.3). A rate of 9.3% equates to more than one bone per radiograph, which is cause for concern because it would require intervention by a skilled operator and this would remove some of the objectivity of the method. Fur-

Table A.1 Comparison of the differences between the CASAS bone age assessment and the manual TW method in healthy children, and in children with some pathological conditions. The six bone method (6b) refers to a limited version of the TW RUS method and is an option on the CASAS system.

Mean difference between CASAS and manual TW results	[Teun96]	[Fris96]	[Dray94]
Healthy children	0.15 years (6b)	0 years	Healthy children from the TW 'Golden Series': 64% no stage difference; 32% one stage difference; 4% two stage differences
Turner's Syndrome	0.04 years	0.4 years	
Familial short stature		0.4 years	
Constitutionally tall stature	0.31 years (6b)		

thermore, not shown in Table A.3 are the results by Albanese *et al.* who reported that in children with general growth delay the manual insertion rate was 29%, and for children with skeletal dysplasia the rate increased to 63% [Alba95]. This suggests a significant deficiency in either the performance of the CASAS system, or the training of the operators using the system.

Table A.2 Comparison of the repeatability between the CASAS system and the manual TW method. The six bone method (6b) refers to a limited version of the TW RUS method and is an option on the CASAS system.

Repeatability of the CASAS bone stages	[Tann94c]	[Tann94b]	[Teun96]	[Dray94]	[Schw98]
No stage difference	95%	97%	88%	85%	Difference of 0–0.56 years (median 0.26 years)
One stage difference	5%	3%	10%	14%	
Two stage differences			2%	1%	
Repeatability of the manual TW bone stages					
No stage difference	83%	85%	90%		
One stage difference	17%	15%	10%		
Two stage differences					

Table A.3 The number of manual insertions required with the CASAS system - bones for which the CASAS system required the user to score the bone manually. The six bone method (6b) refers to a limited version of the TW RUS method and is an option on the CASAS system.

Manual insertions	[Tann94a]	[Teun96]	[Fris96]	[Schw98]
Healthy children	5–10%	8%		
Turner's Syndrome		8% (6b)	12.4%	8.7%
Growth hormone deficiency			8.5%	
Familial short stature			9%	
Constitutionally tall stature		8% boys, 14% girls		

Appendix B

Measurement of the pixel intensity and noise characteristics of the Kodak computed radiography system

Experiments were performed to investigate the pixel intensity and noise characteristics of the Kodak Ektascan HR plates processed on the Kodak CR 900 plate reader, with a focus on the characteristics that relate to hand-wrist radiography. All of the experiments were performed using a Philips Optimus generator and Philips SRO2550 x-ray tube. Dose and kilovoltage (kVp) measurements were taken using a Unfors solid-state dosimeter in the 'RF high' mode. The x-ray tube was warmed-up before measurements were started, and both fine and broad focus were used depending on the measurement made. All measurements were made using a single 18 cm x 24 cm cassette with a Kodak HR plate that was chosen to have performance typical of those used in the Christchurch Hospital Radiology Department. The delay between exposure of the CR plate and processing in the CR plate reader was 1 minute. All equipment was under a quality control programme to ensure compliance with the National Radiation Laboratory (Ministry of Health, New Zealand) NRL C5 Code of Safe Practice for the Use of X-rays in Medical Diagnosis (http://www.nrl.moh.govt.nz/regulatory/c5.pdf) (accessed October 2009). However, the linearity and reproducibility of the x-ray tube output were both checked as part of the experimentation (both were very good and the results have not been reproduced here).

Three experiments were performed to investigate the pixel intensity and noise characteristics of the CR system. The first investigated the variation of pixel intensity with radiation dose to the CR plate – the dose response of the CR system. In the second experiment an aluminium stepwedge was exposed at a distance to measure the radiographic noise at var-

ious pixel intensities. In the last set of experiments a phantom hand was used to simulate the attenuation properties of a real hand so that the CR system would process the image as it would for a hand-wrist radiograph. One of the reasons for this was to determine the impact of overlapping bones on the image pixel intensities.

Image pixel intensity and noise statistics were collected using the ImageJ package [Rasb09] and fixed regions of interest.

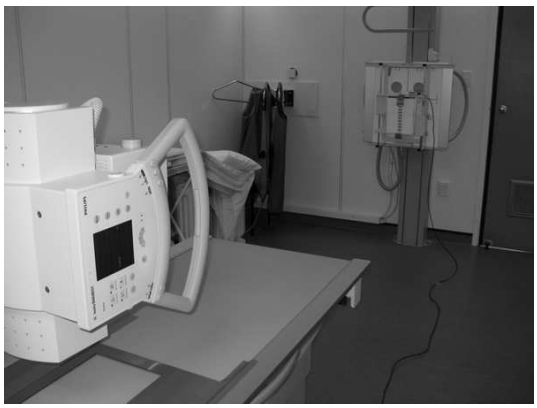
B.1 Dose response of the Kodak CR system

The dose response of the Kodak CR system is the relationship between the radiation dose entering the CR plate, and the output pixel values or intensities. The dose response was measured for 52 kVp, and for 70 kVp with 1 mm of added copper filtration. The dose response was measured at a focus to cassette distance of 200 cm so that the central area of the image was effectively of uniform exposure, having a nonuniformity of 0.13% over a 5x5 cm region. All measurements were made using the Kodak CR setting of 'Pattern', corresponding to minimal post-processing of the image after reading of the plate. This mode is referred to in this thesis as the raw mode.

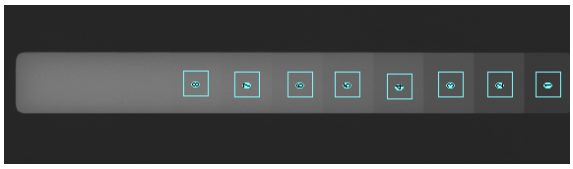
B.2 Pixel intensity and noise characteristics

To investigate the noise characteristics for varying pixel intensity within the image, an aluminium stepwedge was used to generate a range of pixel intensities within a single radiograph. The stepwedge had step sizes of 2.9 ($\pm$0.15) mm. Radiographs were taken of the stepwedge for a range of x-ray exposures all at 52 kVp – the x-ray generator kilovoltage used locally for hand-wrist radiographs. The stepwedge, CR cassette and dosimeter were positioned in a vertical chest stand with a focus to cassette distance of 300 cm (see Figure B.1(a)). This distance was used to reduce nonuniformity due to spherical aberrations from the x-ray focus acting as a point source, and nonuniformities due to the heel effect (Section 5.2.2). All measurements were made using the raw mode of the CR system.

Mean pixel intensity and noise standard deviation measurements were made using 100x100 pixel regions of interest positioned in the centre of each of the eight steps (as illustrated in Figure B.1(b)). The x-ray beam penetration was insufficient in the remaining steps in the stepwedge to make reliable measurements. The stepwedge was x-rayed using differing exposures to give a range of pixel intensities and noise measurements. There were some dust specks on the radiographs, but these were not removed using outlier filtering because the 100x100 pixel region meant their impact on the measurements was negligible. However, an effort was made to position regions of interest to avoid such artefacts.



(a) Stepwedge radiographic exposure at 3 m from the x-ray tube focus (left) to the front of the CR cassette (in the vertical stand). The solid-state dosimeter was placed on top of the vertical CR cassette.



(b) Regions of interest used for analysis of the stepwedge radiographs.

Figure B.1 Experimental setup and analysis of the aluminium stepwedge used for investigation of the computed radiography pixel intensity and noise characteristics.

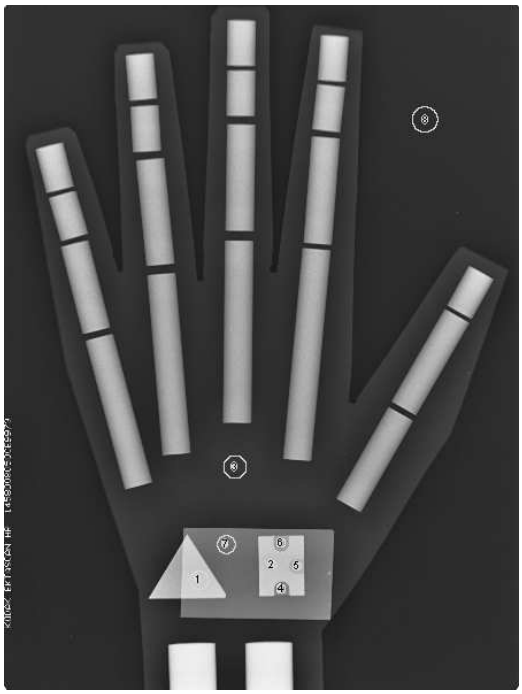
B.3 Hand phantom

A hand phantom was used to simulate human anatomy so that the noise and pixel intensities of the Kodak CR images could be studied without having to expose children's hands to multiple x-rays. By designing areas of homogeneity within the phantom (see Appendix C) it was possible to determine the relationship between the x-ray factors and the noise in certain parts of the image.

The hand phantom was x-rayed using the same x-ray generator voltage as is locally used clinically for hand-wrist radiography (52 kVp), and a x-ray tube focus to CR plate distance of 100 cm. The experimental configuration is illustrated in Figure B.2, along with the ImageJ regions of interest used in the analysis.

All measurements were made using the x-ray tube fine focus and the images were processed using the Kodak CR body part of "Hand AP" and the low exposure optimisation, edge enhancement, and 'EVP' modes enabled; this was equivalent to clinical use of the system.

(a)



(b)

Figure B.2 Experimental setup and analysis of the hand phantom images. (a) The radiographs were taken using a focus-to-cassette distance of 100 cm, x-ray tube light beam diaphragm centred on the cassette, and an Unfors solid-state dosimeter at the top of the hand (anode end of the x-ray tube). The cassette was realigned before each exposure by using a thin perspex template attached to the x-ray table. (b) Radiograph of the phantom showing the regions of interest used for analysis of the pixel intensities and noise. This radiograph shows the aluminium plate used to simulate overlapping carpal bones.

Appendix C

Hand-wrist phantom

A hand phantom was designed and developed by the author[1] to simulate the hand of a young adult undergoing radiography for bone age assessment. The phantom was necessary because the Kodak computed radiography system includes anatomical-based image processing that is only enabled when the user selects the body part being examined. In the case of the hand-wrist radiograph, the user selects the body part "Hand AP" on the Kodak CR 900 system and image processing such as tone scaling, and edge enhancement is performed based on both the image data and preset parameters [Bogu95].

The purpose of the phantom design was to simulate the attenuation properties of the hand but without the structural complexities of bones. A simple geometry was used so that uniform areas were present in the radiograph to allow pixel intensity measurements to be made without the complication of bone structure. The materials consisted of perspex to simulate the soft-tissue of the hand, and aluminium to simulate bone. The x-ray spectrum simulation program Spectra (National Radiation Laboratory, New Zealand) was used to calculate the attenuation properties of perspex and aluminium for an x-ray spectrum based on an x-ray exposure using 52 kVp at 100 cm focus-to-cassette distance and a simulated 3 mm aluminium inherent filtration (this being typical of general purpose x-ray tubes). Under these x-rays conditions it was found that soft tissue is approximately equivalent to perspex of the same thickness. A constant soft tissue thickness of 12 mm was simulated using two sheets of 6 mm perspex glued together using chloroform. No attempt was made to simulate the wedge-shaped soft-tissue thickness of the hand because of the desire to have uniform areas within the radiograph. A cortical bone thickness of 13 mm with an overlying 12 mm of soft tissue equated to 10 mm of aluminium and 12 mm of perspex combined. The equivalence of these simulation materials is only approximate because the calculations

[1]The author would like to thank Grant Wylie, Mechanical Workshop, Medical Physics and Bioengineering Department, Christchurch Hospital, for assistance with machining parts of the phantom.

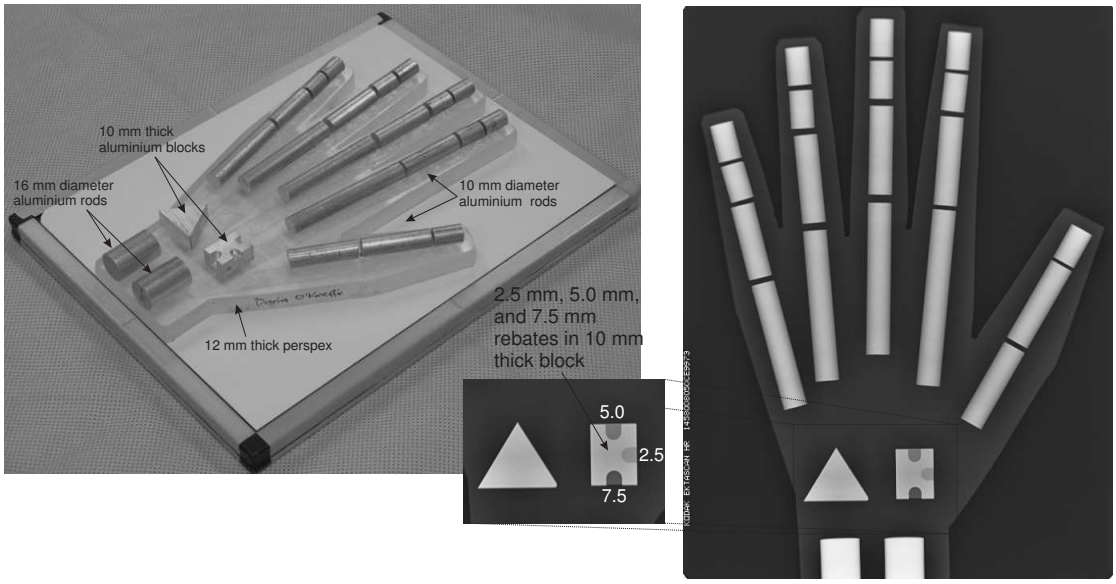


Figure C.1 Details of the phantom used to simulate a hand for hand-wrist radiography. (a) The phantom positioned on a Kodak CR cassette. The phantom was attached to a cardboard template to allow repositioning on the cassette entrance face; (b) Radiograph of the phantom with an enlarged section showing the objects used to simulate edges and contrast of bones in the carpus. This radiograph has very low noise because it was taken using a large x-ray exposure (20 mAs).

were based solely on the x-ray radiation dose at the entrance to the CR plate and did not compensate for changes in performance of the CR plate due to energy spectrum changes caused by the simulation material.

Details of the phantom construction and a radiographic image of the phantom are illustrated in Figure C.1. The carpal bones were simulated using a triangular block to give sharp bone contour angles, and a square block with rebates to give different bone contrasts. Overlapping carpal bones were simulated using a rectangular sheet of 5 mm of aluminium positioned over the blocks.

Appendix D

Bone contour marking and manual data extraction programs

Two MATLAB® programs were developed to manually extract data from the hand-wrist radiographs. Both were developed using the GUIDE graphical user interface tool in MATLAB®.

The first program was designed to make it easy to mark carpal bone contours on the radiographs. After opening the image as a DICOM file (although other common image formats are supported), the user was presented with an interface for marking the eight carpal bones: pisiform, triquetral, hamate, capitate, lunate, scaphoid, trapezoid, and trapezium. This interface included the facility to zoom and pan the image, as well as change image display window level and width (Figure D.1). The bone contours were entered as a collection of points, with the ability to back-step the entries and to flag the confidence in the part of the contour being entered. This confidence entry allowed the user to enter closed contours even if they were not absolutely sure about part of the contour because of poor edge contrast, or complex overlapping structures. At the press of a single key, the interface would toggle between display of the current contour and display of all contours that had been entered up to that point. This improved the ability to resolve complex overlapping structures. The contours were stored as MATLAB® mat-files containing contour point data and DICOM file-header information. The x-y contour point pairs included an integer of either '1' for high confidence edges, or '2' for low confidence edges.

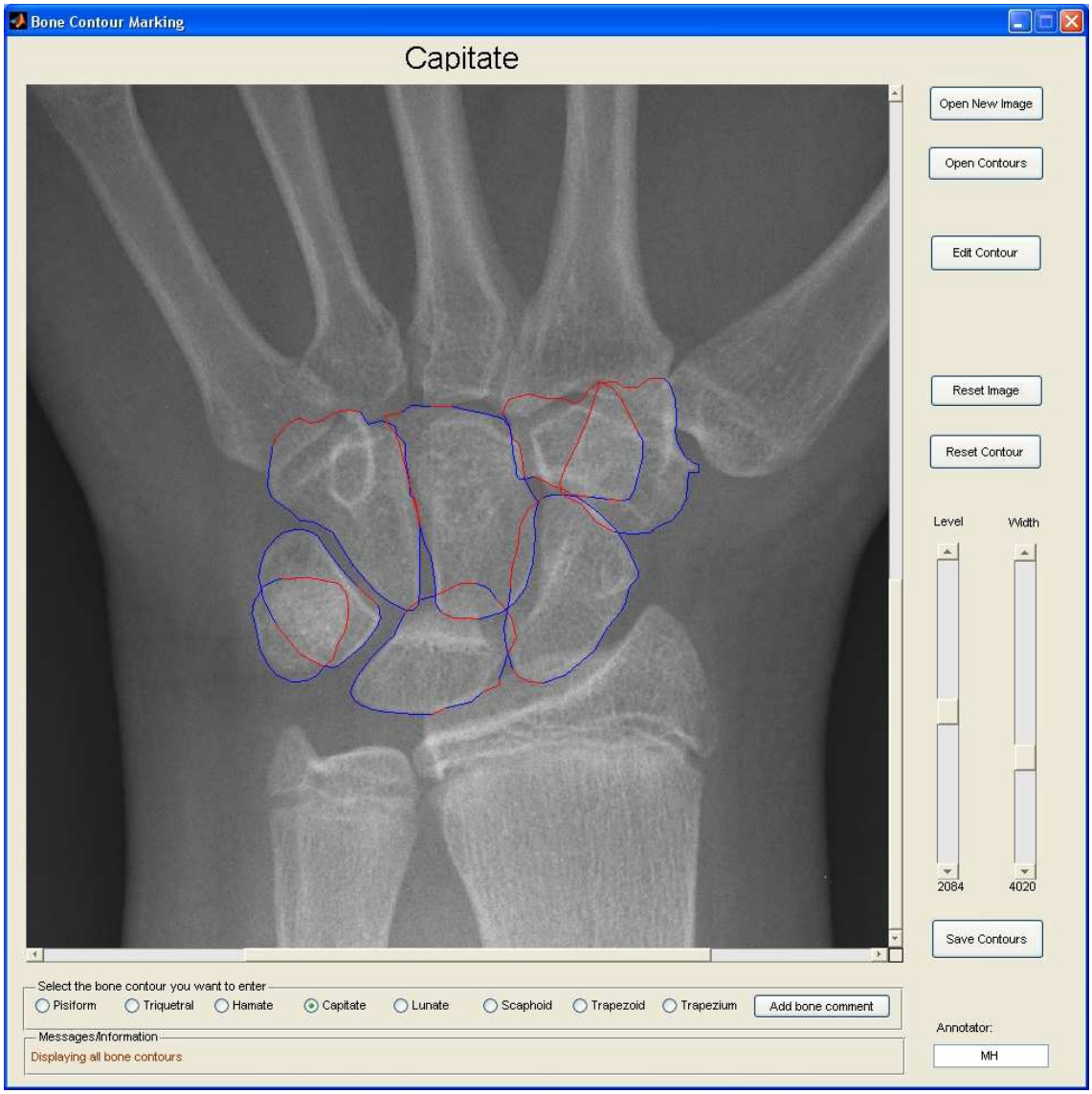


Figure D.1 The bone contour marking program allows the user to open DICOM images, zoom, pan, and window the image, and mark the eight carpal bones. The contour marking included the facility for the user to either label the part of the bone contour being entered as of high confidence (shown here in blue) or low confidence (shown here in red). The contours shown in this figure are straight lines between the marked contour points.

The second program was used to manually delineate the radiographic border, the midline of the third metacarpal, midline of the first phalanx, and the radius-ulna junction—the point at which the radius and ulna internal edges do or would converge if extrapolated. This program used a similar interface to the first program, including the facility to zoom, pan, and window the images.

Figure D.2 The manual extraction program allowed manual marking of the radiation field, the midline of the third metacarpal, the midline of the first phalanx, and the radius-ulna junction. These markings are shown in blue.

Appendix E

Development dataset

A development dataset was used to investigate image denoising and segmentation algorithms. The dataset consisted of eight hand-wrist radiographs: four with non-overlapping carpal bones, and four with overlapping carpal bones (Figure E.1). These eight radiographs were chosen by a radiologist who had extensive experience using the Greulich and Pyle bone age assessment method. The radiographs were chosen to be typical of bone development across a range of maturity from young child to young adult, and avoiding any cases with bone growth disorders. The chronological age range for the radiographs was 22 month female to 16 year male. All radiographs used Kodak Ektascan HR plates processed in a Kodak CR 900 plate reader. They were were exported as DICOM images from the Christchurch Hospital PACS system. (See Section 2.3.5 for details of ethics committee permission to use these images)

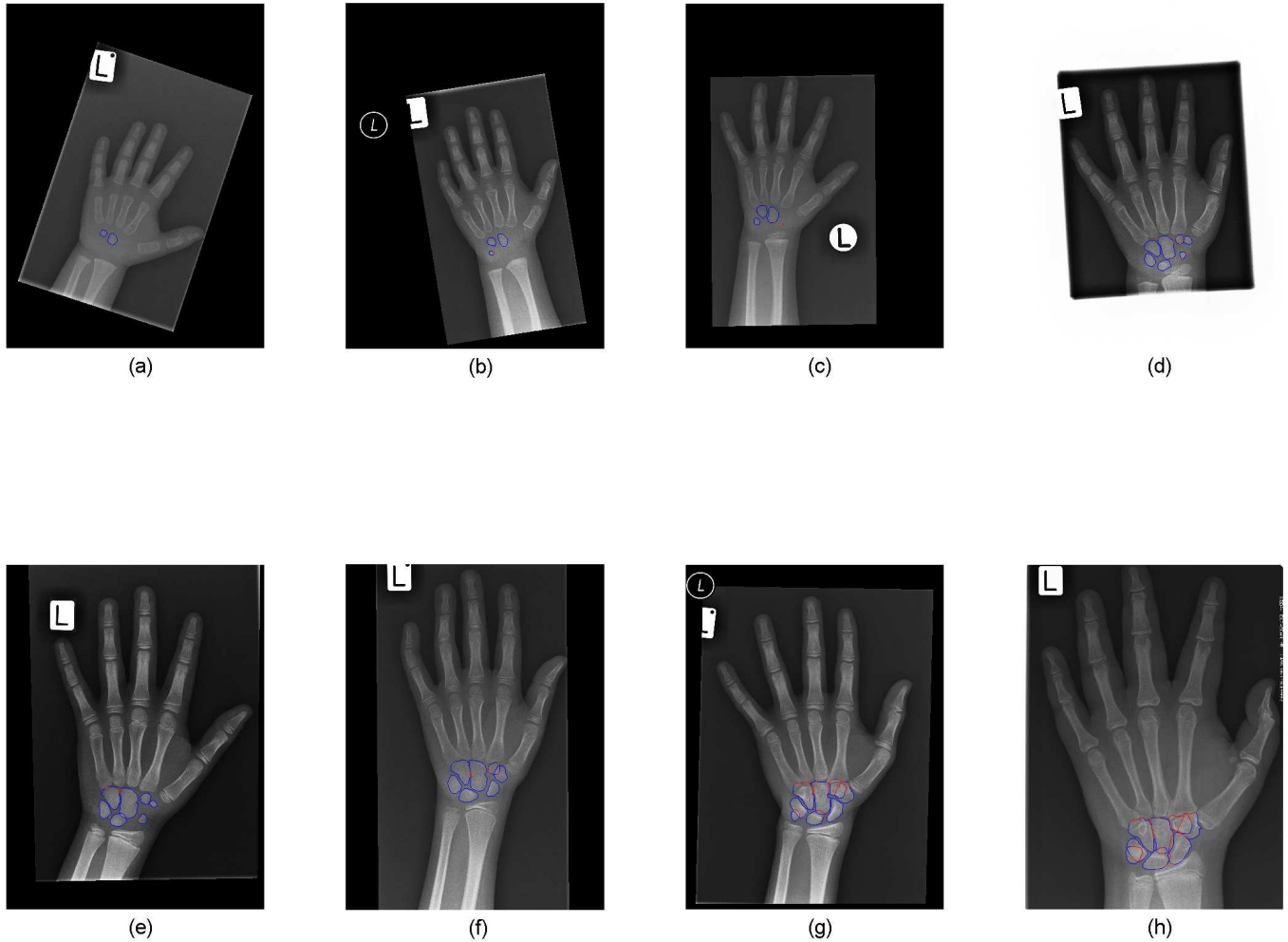


Figure E.1 The eight hand-wrist radiographs used as the development dataset for both denoising and carpal bone segmentation. The original radiographs are shown with contours marked by a radiologist. The contours were colour-coded according to the confidence the radiologist had in the bone edge - blue was high confidence; red was low confidence. (a)–(d) non-overlapping carpal bones; (e)–(h) overlapping carpal bones.

Appendix F

Glossary of terms

Affine transformation - an linear spatial transformation translation that maps straight lines to straight lines by a combination of shearing, rotation, or scaling, followed by a shift translation.

Calcification – the accumulation of calcium or calcium salts in tissue.

Cross-sectional data – data collected from observations or measurements of subjects at a single point in time.

Deleted neighbourhood –A set of pixels about a central pixel that omits the central pixel itself. For a central pixel intensity a and neighbourhood size set by δ the deleted neighbourhood is the set $\{x : 0 < |x - a| < \delta\} - \{a\}$

Diaphysis – the shaft of a long bone.

Differential diagnosis – a method of distinguishing a disease from other diseases that have similar symptoms.

Distal – more distant from a reference point or the midline of the body.

Epiphysis – a section of cartilage or bone that is separated from the main bone by a growth plate.

Growth plate fusion – joining together of the epiphysis and metaphysis of a bone.

Linear filtering – the application of a linear operators to a signal. These filters can usually be implemented using convolution operations.

Longitudinal data – data collected from a study involving repeated observations or measurements of one or more variables collected over a period of time.

Metaphysis – the end of the main bone nearest the growth plate.

Morphology – the study of shape, form, or structure.

Phalanges – the digits of the hand.

Physis – the growth area of a bone (growth plate).

Proximal – closest to a reference point or the midline of the body.

Radiographic mottle – The fluctuation in pixel intensities between regions of an x-ray image due to system noise [Barn82].

Sesamoid – a small mass of bone that forms in a tendon.

References

[Alba95] A. Albanese, C. Hall, and R. Stanhope, "The use of a computerized method of bone age assessment in clinical practice," *Hormone Research*, vol. 44 (suppl 3), pp. 2–7, 1995.

[Ande54] T. W. Anderson and D. A. Darling, "A test of goodness of fit," *Journal of the American Statistical Association*, vol. 49, no. 268, pp. 765–769, 1954.

[Aria09] E. Arias-Castro and D. L. Donoho, "Does median filtering truly preserve edges better than linear filtering?," *Annals of Statistics*, vol. 37, no. 3, pp. 1172–1206, 2009.

[Aube06] G. Aubert and P. Kornprobst, *Mathematical problems in image processing - partial differential equations and the calculus of variations*, Applied Mathematical Series. Springer, 2nd edition, 2006.

[Barn82] G. T. Barnes, "Radiographic mottle: A comprehensive theory," *Medical Physics*, vol. 9, no. 5, pp. 656–667, 1982.

[Barr04] H. H. Barrett and K. J. Myers, *Foundations of image science*, Wiley series in pure and applied optics. Wiley-Interscience, Hoboken, NJ, 2004.

[Behi01] G. Behiels, F. Maes, D. Vandermeulen, and P. Suetens, "Retrospective heel effect correction in conventional radiography," in *Proc. IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA)*. 2001, pp. 87–94, IEEE Press.

[Behi02] G. Behiels, F. Maes, D. Vandermeulen, and P. Suetens, "Retrospective correction of the heel effect in hand radiographs," *Medical Image Analysis*, vol. 6, pp. 183–190, 2002.

[Bers01] M. J. Berst, L. Dolan, M. M. Bogdanowicz, M. A. Stevens, S. Chow, and E. A. Brandser, "Effect of knowledge of chronological age on the variability of pediatric bone age determined use the Greulich and Pyle standards," *American Journal of Roentgenology*, vol. 176, pp. 507–510, 2001.

[Blac98] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger, "Robust anisotropic diffusion," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 421–432, 1998.

[Bocc03] L. Bocchi, F. Ferrara, I. Nicoletti, and G. Valli, "An artificial neural network architecture for skeletal age assessment," in *Proc. International Conference on Image Processing (ICIP 2003)*, 2003, vol. 1, pp. 1077–80.

[Bocc04] L. Bocchi and G. Valli, "Self organizing networks for bone identification on hand radiograms," in *Proc. Medicon and Health Telematics 2004*, Ischia, Italy, 2004, AIIMB and IFMBE.

[Bogu95] T. Bogucki, D. P. Trauernicht, and T. E. Kocher, "Characteristics of a storage phosphor system for medical imaging," Technical and scientific monograph no.6, publication no. n-331, Health Sciences Division, Eastman Kodak Co., Rochester, NY, 1995.

[Brox06] T. Brox, J. Weickert, B. Burgeth, and P. Mrázek, "Nonlinear structure tensors," *Image and Vision Computing*, vol. 24, no. 1, pp. 41–55, 2006.

[Buad05] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[Buad08] A. Buades, B. Coll, and J.-M. Morel, "Nonlocal image and movie denoising," *International Journal of Computer Vision*, vol. 76, no. 2, pp. 123–139, 2008.

[Buck83] J. M. H. Buckler, "How to make the most of bone ages," *Archives of Disease in Childhood*, vol. 58, pp. 761–763, 1983.

[Buck00] J. A. Buckwalter, T. A. Einhorn, and S. R. Simon, *Orthopaedic Basic Science*, Amer Acad of Orthopaedic Surgeons, 2nd edition, 2000.

[Bull99] R. K. Bull, P. D. Edwards, P. M. Kemp, S. Fry, and I. A. Hughes, "Bone age assessment: a large scale comparison of the Greulich and Pyle, and Tanner and Whitehouse (TW2) methods," *Archives of Disease in Childhood*, vol. 81, pp. 172–173, 1999.

[Camm09] L. Cammoun, C. A. Castaño-Moraga, E. Muñoz-Moreno, D. Sosa-Cabrera, B. Acar, M. A. Rodriguez-Florido, A. Brun, H. Knutsson, and J. P. Thiran, "A review of tensors and tensor signal processing," in *Tensors in Image Processing and Computer Vision*, pp. 1–32. 2009.

[Cann86] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986.

[Carp93] C. T. Carpenter and E. L. Lester, "Skeletal age determination in young children: analysis of three regions of the hand/wrist film," *Journal of Pediatric Orthopaedics*, vol. 13, no. 1, pp. 779, 1993.

[Chan03] C.-H. Chang, C.-W. Hsieh, T.-L. Jong, and C.-M. Tiu, "A fully automatic computerized bone age assessment procedure based on phalange ossification analysis," in *Proc. 16th IPPR Conference on Computer Vision and Image Processing (CVGIP)*, 2003, pp. 463–468.

[Chan05] T. F. Chan and J. Shen, *Image processing and analysis : variational, PDE, wavelet, and stochastic methods*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2005.

[Clar62] H. H. Clarke and N. R. Hayman, "Reduction of bone age assessments necessary for the skeletal age determination of boys," *Research Quarterly*, vol. 33, no. 2, pp. 202–207, 1962.

[Clin02] National Institute for Clinical Excellence, "Guidance on the use of human growth hormone (somatropin) in children with growth failure," Technology appraisal guidance - no. 42 (www.nice.org.uk), May 2002.

[Cole88] A. J. L. Cole, L. Webb, and T. J. Cole, "Bone age estimation: a comparison of methods," *British Journal of Radiology*, vol. 61, pp. 683–686, 1988.

[Corn03] B. R. Corner, R. M. Narayanan, and S. E. Reichenbach, "Noise estimation in remote sensing imagery using data masking," *International Journal of Remote Sensing*, vol. 24, no. 4, pp. 689–702, 2003.

[Cox94] L. A. Cox, "Preliminary report on the validation of a grammar-based computer system for assessing skeletal maturity with the Tanner-Whitehouse 2 method," *Acta Paeditrica Supplement*, vol. 406, pp. 84–85, 1994.

[Cox96] L. A. Cox, "Tanner-Whitehouse method of assessing skeletal maturity: Problems and common errors," *Hormone Research*, vol. 45 (suppl 2), pp. 53–55, 1996.

[Cox97] L. A. Cox, "The biology of bone maturation and ageing," *Acta Pædiatrica Supplement*, vol. 423, pp. 107–108, 1997.

[Cund88] D. B. Cundall, J. T. Brocklebank, and J. M. H. Buckler, "Which bone age in chronic renal insufficiency and end-stage renal disease," *Pediatric Nephrology*, vol. 2, pp. 200–204, 1988.

[DAgo86] R. B. D'Agostino and M. A. Stephens, Eds., *Goodness-of-fit techniques*, Marcel Dekker, New York, 1986.

[De L99] F. De Luca and J. Baron, "Skeletal maturation," *The Endocrinologist*, vol. 9, pp. 286–293, 1999.

[Dime05] A. Diméglio, Y. P. Charles, J. Daures, V. de Rosa, and B. Kaboré, "Accuracy of the Sauvegrain Method in determining skeletal age during puberty," *Journal of Bone and Joint Surgery*, vol. 87, no. 8, pp. 1689–1696, 2005.

[Dine95] M. S. Dinesh and A. R. Bhanuprakash, "Vision system for bone measurements from digital hand radiograph," in *Proc. RC-IEEE-EMBS 14th BMESI*, 1995, pp. SPC9–SPC10.

[Doi06] K. Doi, "Diagnostic imaging over the last 50 years: research and development in medical imaging science and technology," *Physics in Medicine and Biology*, vol. 51, no. 13, pp. R5–R27, 2006.

[Dray94] N. M. Drayer and L. A. Cox, "Assessment of bone ages by the Tanner-Whitehouse method using a computer-aided system," *Acta Pædiatrica Supplement*, vol. 406, pp. 77–80, 1994.

[Effo94] N. D. Efford, "Knowledge-based segmentation and feature analysis of hand-wrist radiographs," Tech. Rep. 94.31, University of Leeds, October 1994 1994.

[Eklo67] O. Eklöf and H. Ringertz, "A method for assessment of skeletal maturity," *Annales de Radiologie.*, vol. 10, no. 3, pp. 330–6, 1967.

[Fan01] B.-C. Fan, C.-W. Hsieh, T.-L. Jong, and C.-M. Tiu, "Automatic bone age estimation based on carpal-bone image - a preliminary report," *Chinese Medical Journal (Taipei)*, vol. 64, pp. 203–208, 2001.

[Ferz09] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of Just Noticeable Blur (JNB)," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 717–728, 2009.

[Foi08] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1737–1754, 2008.

[Fris96] H. Frisch, S. Riedl, and T. Waldhör, "Computer-aided estimation of skeletal age and comparison with bone age evaluations by the method of Greulich-Pyle and Tanner-Whitehouse," *Pediatric Radiology*, vol. 26, pp. 226–231, 1996.

[Fros09] I. Frosio and N. A. Borghese, "Statistical based impulsive noise removal in digital radiography," *IEEE Transactions on Medical Imaging*, vol. 28, no. 1, pp. 3–16, 2009.

[Gabo97] R. S. Gaborski and T. F. Pawlicki, "Apparatus and method for identifying specific bone regions on digital x-ray images (US Patent 5,696,805)," 1997.

[Gall10] J. Gallet, "The most powerful diagnostic image processing software from Carestream Health - DirectView EVP Plus Software," Technical Brief Series available from http://www.carestreamhealth.com/awp-cr.html (accessed July 2010), 2010.

[Garn60] S. M. Garn and C. G. Rohmann, "Variability in the order of ossification of the bony centers of the hand and wrist," *American Journal of Physical Anthropology*, vol. 18, pp. 219–30, 1960.

[Garn05] R. Garnett, T. Huegerich, C. Chui, and W. He, "A universal noise removal algorithm with an impulse detector," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1747–1754, 2005.

[Gert07a] A. Gertych, E. Pietka, and B. Liu, "Segmentation of regions of interest and post-segmentation edge location improvement in computer-aided bone age assessment," *Pattern Analysis and Applications*, vol. 10, no. 2, pp. 115–123, 2007.

[Gert07b] A. Gertych, A. Zhang, J. Sayre, S. Pospiech-Kurkowska, and H. K. Huang, "Bone age assessment of children using a digital hand atlas," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 322–331, 2007.

[Gior07] D. Giordano, R. Leonardi, F. Maiorana, G. Scarciofalo, and C. Spampinato, "Epiphysis and metaphysis extraction and classification by adaptive thresholding and DoG filtering for automated skeletal bone age analysis," in *Proc. 29th Annual International Conference of IEEE-EMBS, Engineering in Medicine and Biology Society (EMBC'07)*, 2007, pp. 6551–6556.

[Grah72] C. B. Graham, "Assessment of bone maturation - methods and pitfalls," *Radiologic Clinics of North America*, vol. X, no. 2, pp. 185–202, 1972.

[Grav04] P. Gravel, G. Beaudoin, and J. A. De Guise, "A method for modeling noise in medical images," *IEEE Transactions on Medical Imaging*, vol. 23, no. 10, pp. 1221–1232, 2004.

[Gree01] F. S. Greenspan and D. G. Gardner, *Basic and Clinical Endocrinology*, Lange Medical Books/McCraw-Hill, New York, 6th edition, 2001.

[Greu59] W. W. Greulich and S. I. Pyle, *Radiographic atlas of skeletal development of the hand and wrist*, Stanford University Press, Stanford, California, 2nd edition, 1959.

[Gros95] G. W. Gross, J. M. Bonne, and D. M. Bishop, "Pediatric skeletal age: determination with neural networks," *Radiology*, vol. 195, pp. 689–695, 1995.

[Hait06] F. Haiter-Neto, L. M. Kurita, A. V. Menezes, and M. S. Casanova, "Skeletal age assessment: A comparison of 3 methods," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 130, no. 4, 2006.

[Heal05] Commonwealth Department of Health and Ageing, "Guidelines for the availability of human growth hormone (hGH) as a pharmaceutical benefit," Tech. Rep., April 2005.

[Hein06] J. J. Heine and M. Behera, "Aspects of signal-dependent noise characterization," *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, vol. 23, no. 4, pp. 806–815, 2006.

[Hend02] W. R. Hendee and E. R. Ritenour, *Medical Imaging Physics*, John Wiley and Sons, Inc., New York, 4th edition, 2002.

[Hill94a] K. Hill, "Image processing system and method for automatic feature extraction (European Patent Application 94303075.9, withdrawn 1997)," 1994.

[Hill94b] K. Hill and P. B. Pynsent, "A fully automated bone-ageing system," *Acta Paeditrica Supplement*, vol. 406, pp. 81–83, 1994.

[Hsie07a] C.-W. Hsieh, T.-L. Jong, Y.-H. Chou, and C.-M. Tiu, "Computerized geometric features of carpal bone for bone age estimation," *Chinese Medical Journal*, vol. 120, no. 9, pp. 767–770, 2007.

[Hsie07b] C.-W. Hsieh, T.-L. Jong, and C.-M. Tiu, "Bone age estimation based on phalanx information with fuzzy constrain of carpals," *Medical and Biological Engineering and Computing*, vol. 45, no. 3, pp. 283–295, 2007.

[Hsie08a] C.-W. Hsieh, T.-L. Jong, and C.-M. Tiu, "Carpal growth assessment based on fuzzy description," in *Proc. IEEE Conference on Soft Computing on Industrial Applications (SMCia/08)*, 2008, pp. 355–358.

[Hsie08b] C.-W. Hsieh, T.-C. Liu, T.-L. Jong, and C.-M. Tiu, "Turner's Syndrome screening based on hand image assessment," in *Proc. International Conference on Technology and Applications in Biomedicine (ITAB 2008)*, 2008, pp. 378–379.

[Imme96] J. Immerkær, "Fast noise variance estimation," *Computer Vision and Image Understanding*, vol. 64, no. 2, pp. 300–302, 1996.

[Jahn02] B. Jähne, *Digital image processing*, Springer, Berlin ; New York, 5th rev. and extended edition, 2002.

[John65] F. E. Johnston and S. B. Jahina, "The contribution of the carpal bones to the assessment of skeletal age," *American Journal of Physical Anthropology*, vol. 23, no. 4, pp. 349–354, 1965.

[John83] H. E. Johns and J. R. Cunningham, *The physics of radiology*, Charles C. Thomas, Springfield, Ill., U.S.A., 4th edition, 1983.

[Kapl82] S. Kaplan, *Clinical Pediatric and Adolescent Endocrinology*, W.B. Saunders, Philadelphia, 1982.

[Kauf99] F. R. Kaufman and J. P. Sy, "Regular monitoring of bone age is useful in children treated with growth hormone," *Pediatrics*, vol. 104, pp. 1039–1042, 1999.

[Kemp99] S. F. Kemp and J. P. Sy, "Analysis of bone age data from National Cooperative Growth Study Substudy VII," *Pediatrics*, vol. 104, pp. 1031–1036, 1999.

[King94] D. G. King, D. M. Stevenson, M. P. O'Sullivan, A. M. Cook, V. P. L. Hornsby, I. G. Jefferson, and P. R. King, "Reproducibility of bone ages when performed by radiology registrars: an audit of Tanner and Whitehouse II versus Greulich and Pyle methods," *British Journal of Radiology*, vol. 67, pp. 848–851, 1994.

[Kirk98] D. R. Kirks and N. T. Griscom, *Practical Pediatric Imaging: Diagnositic Radiology of Infants and Children*, Lippincott-Raven Publishers, 3rd edition, 1998.

[Knol00] G. F. Knoll, *Radiation detection and measurement*, Wiley, New York, 3rd edition, 2000.

[Ko95] C.-C. Ko, C.-W. Mao, C.-J. Lin, and Y.-N. Sun, "Image analysis for skeletal evaluation of carpal bones," in *Proc. Visual Communications and Image Processing*, Taipei, 1995, vol. 2501, pp. 951–961, SPIE.

[Kuan85] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-7, no. 2, pp. 165–177, 1985.

[Kwab85] S. A. Kwabwes, S. K. Pal, and R. A. King, "Recognition of bones from x-rays of the hand," *International Journal of Systems Science*, vol. 16, no. 4, pp. 403–413, 1985.

[Lee80] J.-S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-2, no. 2, pp. 165–168, 1980.

[Lee81] J.-S. Lee, "Refined filtering of image noise using local statistics," *Computer Graphics and Image Processing*, vol. 15, no. 4, pp. 380–389, 1981.

[Lee08] Jong-Min Lee and Whoi-Yul Kim, "Epiphyses extraction method using shape information for left hand radiography," in *Proc. International Conference on Convergence and Hybrid Information Technology (ICHIT '08)*, 2008, pp. 319–326.

[Leit87] H. R. Leite, M. T. O'Reilly, and J. M. Close, "Skeletal age assessment using the first, second, and third fingers of the hand," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 92, pp. 492–498, 1987.

[Levi72] E. Levine, "Contributions of carpal bones and the epiphyseal centres of the hand to the assessment of skeletal maturity," *Human Biology*, vol. 44, no. 2, pp. 317–327, 1972.

[Lin04] P. Lin, F. Zhang, Y. Yang, and C.-X. Zheng, "Carpal-bone feature extraction analysis in skeletal age assessment based on deformable model," *Journal of Computer Science and Technology*, vol. 4, no. 3, pp. 152–156, 2004.

[Lin05] P. Lin, C. Zheng, F. Zhang, and Y. Yang, "X-ray carpal-bone image boundary feature analysis using region statistical feature based level set method for skeletal age assessment application," *Optica Applicata*, vol. 35, no. 2, pp. 283–294, 2005.

[Liu08] J. Liu, J. Qi, Z. Liu, Q. Ning, and X. Luo, "Automatic bone age assessment based on intelligent algorithms and comparison with TW3 method," *Computerized Medical Imaging and Graphics*, vol. 32, no. 8, pp. 678–684, 2008.

[Luis03] R. de Luis-García, M. Martín-Fernández, J. I. Arribas, and C. Alberola-López, "A fully automatic algorithm for contour detection of bones in hand radiographs using active contours," in *Proc. International Conference on Image Processing*. 2003, vol. 3, pp. 421–424, IEEE.

[Mack07] A. Mackenzie and I. D. Honey, "Characterization of noise sources for two generations of computed radiography systems using powder and crystalline photostimulable phosphors," *Medical Physics*, vol. 34, no. 8, pp. 3345–3357, 2007.

[Mack08] A. Mackenzie, "Validation of correction methods for the non-linear response of digital radiography systems," *British Journal of Radiology*, vol. 81, no. 964, pp. 341–345, 2008.

[Mahm98] S. Mahmoodi, *A knowledge based computer vision system for skeletal age assessment of children*, PhD thesis, University of Newcastle upon Tyne, United Kingdom., 1998.

[Mahm00] S. Mahmoodi, B. S. Sharif, E. G. Chester, J. P. Owen, and R. E. J. Lee, "Skeletal growth estimation using radiographic image processing and analysis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 4, no. 4, pp. 292–297, 2000.

[Mano94] G. K. Manos, A. Y. Cairns, I. W. Rickets, and D. Sinclair, "Segmenting radiographs of the hand and wrist," *Computer Methods and Programs in Biomedicine*, vol. 43, pp. 227–237, 1994.

[Marq01] A. M. Marques Da Silva, S. Delgado Olabarriaga, C. A. Dietrich, and C. A. A. Schmitz, "On determining a signature for skeletal maturity," in *Proc. XIV Brazilian Symposium on Computer Graphics and Image Processing*, 2001, pp. 246–251.

[Mars77] W. A. Marshall, *Human Growth and its Disorders*, Academic Press Inc., London, 1977.

[Mars09] N. W. Marshall, "An examination of automatic exposure control regimes for two digital radiography systems," *Physics in Medicine and Biology*, vol. 54, no. 15, pp. 4645–4670, 2009.

[Mart02] W. L. Martinez and A. R. Martinez, *Computational statistics handbook with Matlab*, Chapman and Hall/CRC, Boca Raton, 2002.

[Mart03] M. Á. Martín-Fernández, M. Martín-Fernández, and C. Alberola-López, "Automatic bone age assessment: a registration approach," in *Proc. SPIE Medical Imaging 2003: Image Processing*. 2003, vol. 5032, pp. 1765–1776, SPIE.

[Mart09] M. Á. Martín-Fernández, R. Cárdenes, E. Muñoz-Moreno, R. d. Luis-García, M. Martín-Fernández, and C. Alberola-López, "Automatic articulated registration of hand radiographs," *Image and Vision Computing*, vol. 27, no. 8, pp. 1207–1222, 2009.

[Mich89] D. J. Michael and A. C. Nelson, "HANDX: A model-based system for automatic segmentation of bones from the digital hand radiograph," *IEEE Transactions on Medical Imaging*, vol. 8, no. 1, pp. 64–69, 1989.

[Moli04] L. Molinari, T. Gasser, and R. H. Largo, "TW3 bone ages: RUS/CB and gender differences of percentiles for score and score increments," *Annals of Human Biology*, vol. 31, no. 4, pp. 421–435, 2004.

[Morr94] D. T. Morris and C. F. Walshaw, "Segmentation of the finger bones as a prerequisite for the determination of bone age," *Image and Vision Computing*, vol. 12, no. 4, pp. 239–245, 1994.

[Mura97] M. Murata, "Population-specific reference values for bone age," *Acta Pædiatrica Supplement*, vol. 423, pp. 113–114, 1997.

[Niem02] M. Niemeijer, *Automating skeletal age assessment*, Masters thesis, Universiteit Utrecht, 2002.

[Niko04] M. Nikolova, "A variational approach to remove outliers and impulse noise," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1, pp. 99–120, 2004.

[OKee04] D. S. O'Keeffe and R. W. McLeod, "Computed radiography as a gamma ray detector - dose response and applications," *Physics in Medicine and Biology*, vol. 49, no. 16, pp. 3559–3572, 2004.

[Olse93] S. I. Olsen, "Estimation of noise in images: an evaluation," *CVGIP: Graphical Models and Image Processing*, vol. 55, no. 4, pp. 319–323, 1993.

[Onte96] F. K. Ontell, M. Ivanovic, D. S. Ablin, and T. W. Barlow, "Bone age in children of diverse ethnicity," *American Journal of Roentgenology*, vol. 167, pp. 1395–1398, 1996.

[Pal83] S. K. Pal and R. A. King, "On edge detection of x-rays using fuzzy sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, no. 1, pp. 69–77, 1983.

[Pero90] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[Piet91] E. Pietka, M. F. McNitt-Gray, M. L. Kuo, and H. K. Huang, "Computer-assisted phalangeal analysis in skeletal age assessment," *IEEE Transactions on Medical Imaging*, vol. 10, no. 4, pp. 616–620, 1991.

[Piet93] E. Pietka, L. Kaabi, M. L. Kuo, and H. K. Huang, "Feature extraction in carpal-bone analysis," *IEEE Transactions on Medical Imaging*, vol. 12, no. 1, pp. 44–49, 1993.

[Piet95] E. Pietka, "Computer-assisted bone age assessment based on features automatically extracted from a hand radiograph," *Computerized Medical Imaging and Graphics*, vol. 19, no. 3, pp. 251–259, 1995.

[Piet97] E. Pietka and H. K. Huang, "Image processing techniques in bone age assessment," in *Medical Imaging Systems - Techniques and Applications: Vol 2 Brain and Skeletal Systems*, C. T. Leondes, Ed., pp. 221–272. Gordan and Brench Science Publishers, 1997.

[Piet01a] E. Pietka, A. Gertych, S. Pospiech, F. Cao, and H. K. Huang, "Computer-assisted bone age assessment: Image preprocessing and epiphyseal/metaphyseal ROI extraction," *IEEE Transactions on Medical Imaging*, vol. 20, no. 8, pp. 715–729, 2001.

[Piet01b] E. Pietka, S. Pospiech, A. Gertych, F. Cao, H. K. Huang, and V. Gilsanz, "Computer automated approach to the extraction of epiphyseal regions in hand radiographs," *Journal of Digital Imaging*, vol. 14, no. 4, pp. 165–172, 2001.

[Piet03a] E. Pietka, "Computer-assisted bone age assessment–database adjustment," *International Congress Series*, vol. 1256, pp. 87–92, 2003.

[Piet03b] E. Pietka, S. Pospiech-Kurkowska, A. Gertych, and F. Cao, "Integration of computer assisted bone age assessment with clinical PACS," *Computerized Medical Imaging and Graphics*, vol. 27, no. 2-3, pp. 217–228, 2003.

[Piet04] E. Pietka, A. Gertych, S. Pospiech-Kurkowska, F. Cao, H. K. Huang, and V. Gilsanz, "Computer-assisted bone age assessment: Graphical user interface for image processing and comparison," *Journal of Digital Imaging*, vol. 17, no. 3, pp. 175–188, 2004.

[Piet05] E. Pietka, A. Gertych, and K. Witko, "Informatics infrastructure of CAD system," *Computerized Medical Imaging and Graphics*, vol. 29, no. 2-3, pp. 157–169, 2005.

[Plud04] P. Pludowski, M. Lebiedowski, and R. S. Lorenc, "Evaluation of the possibility to assess bone age on the basis of DXA derived hand scanspreliminary results," *Osteoporosis International*, vol. V15, no. 4, pp. 317–322, 2004.

[Prat01] W. K. Pratt, *Digital Image Processing*, John Wiley and Sons, Inc., New York, 3rd edition, 2001.

[Pres07] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes : the art of scientific computing*, Cambridge University Press, Cambridge, UK ; New York, 3rd edition, 2007.

[Rasb09] W. Rasband, "ImageJ 1.43e software, available from http://rsb.info.nih.gov/ij (accessed September 2009)," 2009.

[Riou96] O. Rioul, "Spectral algorithm for removing salt and pepper from images," in *Proc. IEEE Digital Signal Processing Workshop*, 1996, pp. 275–278.

[Roch70a] A. F. Roche, G. H. Davila, B. A. Pasternack, and M. J. Walton, "Some factors influencing the replicability of assessments of skeletal maturity (Greulich-Pyle)," *American Journal of Roentgenology, Radium Therapy and Nuclear Medicine*, vol. 109, no. 2, pp. 299–306, 1970.

[Roch70b] A. F. Roche, C. G. Rohmann, N. Y. French, and G. H. Davila, "Effect of training on replicability of assessments of skeletal maturity (Greulich-Pyle)," *American Journal of Roentgenology, Radium Therapy and Nuclear Medicine*, vol. 108, no. 3, pp. 511–515, 1970.

[Roch88] A. F. Roche, W. C. Chumlea, and D. Thissen, *Assessing the Skeletal Maturity of the Hand-Wrist: FELS Method*, Charles C Thomas, Springfield, Illinois, USA, 1st edition, 1988.

[Roem97] J. N. Roemmich, R. M. Blizzard, S. D. Peddada, R. M. Malina, A. F. Roche, J. M. Tanner, and A. D. Rogol, "Longitudinal assessment of hormonal and physical alterations during normal puberty in boys. IV: Predictions of adult height by the Bayley-Pinneau, Roche-Wainer-Thissen, and Tanner-Whitehouse methods compared," *American Journal of Human Biology*, vol. 9, pp. 371–380, 1997.

[Rous87] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*, Wiley series in probability and mathematical statistics. Applied probability and statistics,. Wiley, New York, 1987.

[Rous93] P. J. Rousseeuw and C. Croux, "Alternatives to the median absolute deviation," *Journal of the American Statistical Association*, vol. 88, no. 424, pp. 1273–1283, 1993.

[Rous02] P. J. Rousseeuw and S. Verboven, "Robust estimation in very small samples," *Computational Statistics and Data Analysis*, vol. 40, no. 4, pp. 741–758, 2002.

[Rowl02] J. A. Rowlands, "The physics of computed radiography," *Physics in Medicine and Biology*, vol. 47, pp. R123–R166, 2002.

[Rucc95] M. Rucci, G. Coppini, I. Nicoletti, D. Chell, and G. Valli, "Automatic analysis of hand radiographs for the assessment of skeletal age: a subsymbolic approach," *Computers and Biomedical Research*, vol. 28, pp. 239–256, 1995.

[Same01] E. Samei, J. A. Seibert, C. E. Willis, M. J. Flynn, E. Mah, and K. L. Junck, "Performance evaluation of computed radiography systems," *Medical Physics*, vol. 28, no. 3, pp. 361–371, 2001.

[Same05] E. Samei, A. Badano, D. Chakraborty, K. Compton, C. Cornelius, K. Corrigan, M. J. Flynn, B. Hemminger, N. Hangiandreou, J. Johnson, M. Moxley, W. Pavlicek, H. Roehrig, L. Rutz, J. Shepard, R. Uzenoff, J. Wang, and C. N. Willis, "Assessment of display performance for medical imaging systems," Tech. Rep., American Association of Physicists in Medicine (AAPM) Task Group 18 (Medical Physics Publishing, Madison, WI). AAPM On-line Report No. 03, 2005.

[Sapi01] G. Sapiro, *Geometric Partial Differential Equations and Image Analysis*, Cambridge University Press, 2001.

[Sato99] K. Sato, K. Ashizawa, M. Anzo, F. Otsuki, S. Kaneko, T. Tanaka, K. Tsukagoshi, A. Nimura, H. Matsuok, N. Matsuo, H. Mitani, and M. Murata, "Setting up an automated system for evaluation of bone age," *Endocrine Journal*, vol. 46, pp. S97–100, 1999.

[Sato01] K. Sato, T. Mito, and H. Mitani, "An accurate method of predicting mandibular growth potential based on bone maturity," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 120, no. 3, pp. 286–293, 2001.

[Schm00] A. Schmeling, W. Reisinger, D. Loreck, K. Vendura, W. Markus, and G. Geserick, "Effects of ethnicity on skeletal maturation: consequences for forensic age estimations," *International Journal of Legal Medicine*, vol. 113, pp. 253–258, 2000.

[Schw98] C. P. Schwarze, D. Arens, H. P. Haber, H. A. Wollmann, G. Biinder, E. I. E. Mayer, and M. B. Ranke, "Bone age in 116 untreated patients with Turner's syndrome rated by a computer-assisted method (CASAS)," *Acta Paediatrica*, vol. 87, pp. 1146–1150, 1998.

[Seib06] J. A. Seibert, T. M. Bogucki, T. Ciona, W. Huda, A. Karellas, J. R. Mercier, E. Samei, S. J. Shepard, B. K. Stewart, K. J. Strauss, O. Suleiman, D. Tucker, R. A. Uzenoff, J. C. Weiser, and C. E. Willis, "Acceptance testing and quality control of photostimulable storage phosphor imaging systems," report available from http://www.aapm.org/pubs/reports/rpt_93.pdf (accessed October 2009), 2006.

[Shim97] H. Shim, B. J. Liu, R. K. Taira, and T. R. Hall, "Object-oriented approach to the automatic segmentation of bones from pediatric hand radiographs," in *Proc. Medical Imaging 1997: Image Processing*, K. M. Hanson, Ed., Newport Beach, CA, USA, 1997, vol. 3034, pp. 95–105, SPIE.

[Soci00] GH Research Society, "Consensus guidelines for the diagnosis and treatment of growth hormone (GH) deficiency in childhood and adolescence: summary statement of the GH Research Society," *Journal of Clinical Endocrinology and Metabolism*, 2000.

[Spen92] A. P. Spence and E. B. Mason, *Human Anatomy and Physiology*, West Publishing Company, 4th edition, 1992.

[Sun94] Y. N. Sun, C. C. Ko, C. W. Mao, and C. J. Lin, "A computer system for skeletal growth measurement," *Computers and Biomedical Research*, vol. 27, pp. 2–12, 1994.

[Tai08] Shen-Chuan Tai and Shih-Ming Yang, "A fast method for image noise estimation using Laplacian operator and adaptive edge detection," in *Proc. 3rd International Symposium on Communications, Control and Signal Processing (ISCCSP 2008)*, 2008, pp. 1077–1081.

[Tann83] J. M. Tanner, R. H. Whitehouse, N. Cameron, W. A. Marshall, M. J. R. Healy, and H. Goldstein, *Assessment of skeletal maturity and prediction of adult height (TW2 method)*, Academic Press, London, 2nd edition, 1983.

[Tann94a] J. M. Tanner and R. D. Gibbons, "Automatic bone age measurement using computerized image analysis," *Journal of Pediatric Endocrinology*, vol. 7, no. 2, pp. 141–145, 1994.

[Tann94b] J. M. Tanner and R. D. Gibbons, "A computerized image analysis system for estimating Tanner-Whitehouse 2 bone age," *Hormone Research*, vol. 42, pp. 282–287, 1994.

[Tann94c] J. M. Tanner, D. Oshman, G. Lindgren, J. A. Grunbaum, R. Elsouki, and D. Labarthe, "Reliability and validity of computer-assisted estimates of Tanner-Whitehouse skeletal maturity (CASAS) - comparison with the manual method," *Hormone Research*, vol. 42, no. 6, pp. 288–294, 1994.

[Tann01] J. M. Tanner, M. J. R. Healy, H. Goldstein, and N. Cameron, *Assessment of skeletal maturity and prediction of adult height (TW3 method)*, WB Saunders, 3rd edition, 2001.

[Tara76] J. Taranger, B. Bruning, I. Claesson, P. Karlberg, T. Landström, and B. Lindström, "(IV.) Skeletal development from birth to 7 years," *Acta Pædiatrica Scandinavica Supplement*, vol. 258, pp. 98–108, 1976.

[Teun96] A. van Teunenbroek, W. de Waal, A. Roks, P. Chinafo, M. Fokker, P. Mulder, S. M. de Munick Keizer-Schrama, and S. Drop, "Computer-aided skeletal age scores in healthy children, girls with Turner Syndrome, and in children with constitutionally tall stature," *Pediatric Research*, vol. 39, no. 2, pp. 360–367, 1996.

[Thod09] H. H. Thodberg, S. Kreiborg, A. Juul, and K. D. Pedersen, "The BoneXpert method for automated determination of skeletal maturity," *IEEE Transactions on Medical Imaging*, vol. 28, no. 1, pp. 52–66, 2009.

[Tris08] A. Tristán-Vega and J. I. Arribas, "A radius and ulna TW3 bone age assessment system," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 5, pp. 1463–1476, 2008.

[Tsai85] W.-H. Tsai, "Moment-preserving thresholding: A new approach," *Computer Vision, Graphics, and Image Processing*, vol. 29, pp. 377–393, 1985.

[Tsai01] A. Tsai, A. Yezzi, and A. S. Willsky, "Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification," *IEEE Transactions on Image Processing*, vol. 10, no. 8, pp. 1169–1186, 2001.

[Tsao08] S. Tsao, A. Gertych, A. Zhang, B. J. Liu, and H. K. Huang, "Automated bone age assessment of older children using the radius," in *Proc. SPIE- Progress in Biomedical Optics and Imaging*, 2008, vol. 6919.

[Tsch] D. Tschumperlé, "The CImg Library: C++ template image processing toolkit, available from http://cimg.sourceforge.net/ (accessed September 2009)," .

[Tsch02] D. Tschumperlé, *PDE's based regularization of multivalued images and applications*, PhD thesis, Université de Nice - Sophia Antipolis, 2002.

[Tsch05] D. Tschumperlé and R. Deriche, "Vector-valued image regularization with PDEs: a common framework for different applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 506–517, 2005.

[Tsch06a] D. Tschumperlé, "Fast anisotropic smoothing of multi-valued images using curvature-preserving PDE's," *International Journal of Computer Vision*, vol. 68, no. 1, pp. 65–82, 2006.

[Tsch06b] D. Tschumperlé, J. Fadili, and Y. Bentolila, "Wire structure pattern extraction and tracking from X-ray images of composite mechanisms," in *Proc. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006)*, 2006, vol. 2, pp. 2461–2466.

[Vejv81] M. Vejvoda and D. B. Grant, "Discordant bone maturation of the hand in children with precocious puberty and congential adrenal hyperplasia," *Acta Pædiatrica Scandanavia*, vol. 70, pp. 903–905, 1981.

[Voge00] F. Vogelsang, M. Kohnen, H. Schneider, F. Weiler, B. B. Wein, M. Kilbinger, and R. W. Günther, "Skeletal maturity determination from hand radiograph by model based analysis," in *Proc. Medical Imaging 2000: Image Processing*, K. M. Hanson, Ed., San Diego, USA, 2000, vol. Vol. 3979, pp. 294–305, SPIE.

[Wang04] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[Warr00] M. W. Warren, K. R. Smith, P. R. Stubbefield, S. S. Martin, and H. A. Walsh-Haney, "Use of radiographic atlases in mass fatality," *Journal of Forensic Sciences*, vol. 45, no. 2, pp. 467–470, 2000.

[Wee08] Chong-Yaw Wee and R. Paramesran, "Image sharpness measure using eigenvalues," in *Proc. International Conference on Signal Processing Proceedings, ICSP*, 2008, pp. 840–843.

[Weic08] J. Weickert, *Anisotropic diffusion in image processing*, B G Teubner Stuttgart, updated from 1998 edition, 2008.

[Wert96] G. A. Werther, "Growth hormone measurements versus auxology in treatment decisions: The Australian experience," *Journal of Clinical Endocrinology and Metabolism*, vol. 128, pp. S47–51, 1996.

[Wils98] J. D. Wilson, D. W. Foster, H. M. Kronenberg, and P. R. Larsen, *Williams Textbook of Endocrinology*, W B Saunders Company, Philadelphia, 9th edition, 1998.

[Wils99] D. M. Wilson, "Regular monitoring of bone age is not useful in children treated with growth hormone," *Pediatrics*, vol. 104, pp. 1036–1039, 1999.

[Wind01] P. S. Windyga, "Fast impulsive noise removal," *IEEE Transactions on Image Processing*, vol. 10, no. 1, pp. 173–179, 2001.

[Wins05] M. Winslow, X. G. Xu, and B. Yazici, "Development of a simulator for radiographic image optimization," *Computer Methods and Programs in Biomedicine*, vol. 78, no. 3, pp. 179–190, 2005.

[Wyat95] D. Wyatt, D. Mark, and A. Slyper, "Survey of growth hormone treatment practices by 251 pediatric endocrinologists," *Journal of Clinical Endocrinology and Metabolism*, vol. 80, no. 11, pp. 3292–3297, 1995.

[Wyat07] P. Wyatt and H. Nakai, "Developing nonstationary noise estimation for application in edge and corner detection," *IEEE Transactions on Image Processing*, vol. 16, no. 7, pp. 1840–1853, 2007.

[Yaff94] M. J. Yaffe and R. M. Nishikawa, "X-ray imaging concepts: Noise, SNR and DQE," in *Specification, Acceptance Testing and Quality Control of Diagnostic X-ray Imaging Equipment (AAPM Medical Physics Monograph No. 20)*, J. A. Seibert, G. T. Barnes, and R. G. Gould, Eds., pp. 109–144. 1994.

[Yazi07] B. Yazici and S. Yolacan, "A comparison of various tests of normality," *Journal of Statistical Computation and Simulation*, vol. 77, no. 2, pp. 175–183, 2007.

[Yoo04] T. S. Yoo, *Insight into images : principles and practice for segmentation, registration, and image analysis*, A K Peters, Wellesley, Massachusetts, 2004.

[Yoo07] Juwoan Yoo, *Bone age assessment algorithm using normalized shape model and pattern recognition techniques*, M.Sc. thesis, Hanyany University, 2007.

[You96] Y.-L. You, W. Xu, A. Tannenbaum, and M. Kaveh, "Behavioral analysis of anisotropic diffusion in image processing," *IEEE Transactions on Image Processing*, vol. 5, no. 11, pp. 1539–1553, 1996.

[Zach83] M. Zachmann, S. D. Frasier, J. McLaughlin, L. Hurley, and P. Nessi, "Importance and accuracy of bone age ratings in a computerized growth evaluation system," *Hormone Research*, vol. 18, pp. 160–167, 1983.

[Zeri91] J. M. Zerin and R. J. Hernandez, "Approach to skeletal maturation," *Hand Clinics*, vol. 7, no. 1, pp. 53–62, 1991.

[Zhan97] J. Zhang and H. K. Huang, "Automatic background recognition and removal (ABRR) in computed radiography images," *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 762–771, 1997.

[Zhan06] A. Zhang, A. Gertych, S. Kurkowska-Pospiech, B. J. Liu, and H. K. Huang, "Carpal bone analysis in bone age assessment," in *Proc. Medical Imaging 2006: PACS and Imaging Informatics*, S. C. Horii and O. M. Ratib, Eds., San Diego, CA, USA, 2006, vol. 6145, pp. 61450Z–9, SPIE.

[Zhan07] A. Zhang, A. Gertych, and B. J. Liu, "Automatic bone age assessment for young children from newborn to 7-year-old using carpal bones," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 299–310, 2007.

[Zhu09] X. Zhu and P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise," in *Proc. 2009 International Workshop on Quality of Multimedia Experience (QoMEX)*, San Diego, CA, 2009, pp. 64–69 (article number 5246976).